Copy
10/21/99

AD_____

GRANT NUMBER DAMD17-98-1-8042

TITLE: New Classes of Conditional Toxins as Therapeutic Agents Against Breast Cancer

PRINCIPAL INVESTIGATOR: Alexander J. Varshavsky, Ph.D.

CONTRACTING ORGANIZATION: California Institute of Technology
Pasadena, California 91125

REPORT DATE: April 1999

TYPE OF REPORT: Annual

PREPARED FOR:
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. AGENCY USE ONLY *(Leave blank)* | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | April 1999 | Annual (1 Apr 98 - 31 Mar 99) |

**4. TITLE AND SUBTITLE**
New Classes of Conditional Toxins as Therapeutic Agents Against Breast Cancer

**5. FUNDING NUMBERS**
DAMD17-98-1-8042

**6. AUTHOR(S)**
Alexander J. Varshavsky, Ph.D.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
California Institute of Technology
Pasadena, California 91125

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for Public Release; Distribution Unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

During the first year of support by the grant DAMD17-98-1-8042 (The Idea Grant) we focused on the following projects:

1) Development of *sitoxins* (signal-regulated, cleavage-mediated toxins). One problem with the earlier design was a significant toxicity of Cup9p. We addressed this by constructing a mutant of Cup9p that does not bind to DNA, thereby greatly reducing the toxicity of Cup9p. Further work on sitoxins using this mutant is under way.

2) To design better portable degrons (degradation signals) for the *comtoxin* projects (item 3), and also to address the problem described in item 1, we developed a new approach to construct strong degradation signals that consist exclusively of asparagines and lysines. In another advance, we constructed a bivalent inhibitor of the N-end rule pathway.

3) During the first year of this grant, the work on *comtoxins* began with the construction of a DNA library that expresses polypeptide fragments derived from random segments of the proteins of interest. The current state of this ongoing project is described below.

**14. SUBJECT TERMS**
Breast Cancer
proteolysis / degradation signals / codominance / conditional toxins / cancer

**15. NUMBER OF PAGES** 116

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Unlimited |

## FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

__V__ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

__V__ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

__ __ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Alexa Vara 4/20/99

PI - Signature Date

**Table of Contents**

# Progress Report for the DOD Grant DAMD 17-98-1-8042

April 1998-April 1999 (first year of support)

During the first year of support by the grant DAMD17-98-1-8042 (the Idea Grant) we focused on several projects whose current status is described below.

**Development of *sitoxins* (signal-regulated, cleavage-mediated toxins)**

One problem with the earlier design, described in the Idea Grant proposal, was a significant toxicity of Cup9p, which was used as a portable degradation signal (degron). We addressed this problem by constructing a mutant of Cup9p that does not bind to its high-affinity sites in the *S. cerevisiae* DNA, thereby greatly reducing the toxicity of Cup9p. Specifically, a missense mutation was introduced into the alpha-helical region of Cup9p that is known to contact the major groove of DNA and contribute to the high-affinity binding by Cup9p. This mutation retained the alpha-helix of Cup9p, did not interfere with its overall folding, but prevented the high-affinity binding of Cup9p to DNA. The resulting mutant Cup9p was still short-lived *in vivo*, but was no longer toxic, even at high expression levels. Further work on sitoxins using this mutant is under way.

**Construction of degradation signals in the lysine-asparagine sequence space**

Among the problems to be solved on the way to multitarget conditional toxins is the paucity of strong, portable N-degrons that are free of drawbacks of the existing degrons of this class. These drawbacks include the tendency of proteins bearing such degrons to undergo endoproteolytic cleavages that sever the N-degron from the rest of the protein, and thereby contribute to the high background of assays with N-degron-bearing proteins. As a part of our ongoing effort to construct better degradation signals, we developed a new approach, which makes it possible to identify and isolate specific degradation signals in the sequence space of just two amino acids, lysine and asparagine A paper describing these results is being written up. Our findings are briefly summarized below.

The N-degrons, a subset of degradation signals recognized by the N-end rule pathway, comprise a protein's destabilizing N-terminal residue and an internal lysine residue. We showed

that the strength of active N-degron can be markedly increased, without loss of specificity, through the addition of lysine residues. To explore constraints on the structure of N-degrons, we took advantage of a single third-letter difference between the codons for lysine and asparagine. A nearly exhaustive screen was carried out for N-degrons in the lysine-asparagine (K/N) sequence space of the 14-residue K/N regions (16,384 different sequences). 68 of these sequences were found to function as N-degrons, and 3 of them were at least as active as any of the previously known N-degrons. All K/N-based N-degrons lacked the lysine at position 2, and all three of the strongest N-degrons contained lysines at positions 3 and 15. Our findings support the model of the targeting mechanism in which the binding of the E3-E2 complex to the substrate's destabilizing N-terminal residue is followed by a stochastic, time-restricted search for a sterically suitable lysine residue. The strategy of screening a small library that encompasses the entire sequence space of two amino acids should be of use in many settings, including studies of protein targeting and folding. A paper describing these results is being written up.

We plan to employ the lysine-asparagine N-degron discovered in the present work in both the designs of sitoxins above, and in the trans-targeting screens described below.

**Bivalent Inhibitor of Protein Degradation by the N-End Rule Pathway**

In a project relevant to the theme of this grant, we have developed a specific bivalent inhibitor of the N-end rule pathway. The relevance of this advance stems from the fact the bivalent inhibitor (described below) improved the understanding of the targeting (substrate-binding) sites of N-recognin, the E3 component of the ubiquitin-dependent N-end rule pathway, which is at the center of our sitoxins and comtoxin projects. Eventually, the designs of conditional toxins should be able to involve any of the several ubiquitin-dependent proteolytic pathways; the current focus on the N-end rule pathway reflects not only the interest of this laboratory but also the fact that this pathway is particularly well understood mechanistically. Moreover, in staying within the N-end rule pathway, it is possible to use the portable, highly active N-degrons, an advantage that is not available with most of the other ubiquitin-dependent pathways.

The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Ubr1p, the recognition (E3) component of the *Saccharomyces cerevisiae* N-end rule

pathway, contains at least two substrate-binding sites. The type 1 site is specific for N-terminal basic residues Arg, Lys, and His. The type 2 site is specific for N-terminal bulky hydrophobic residues Phe, Leu, Trp, Tyr, and Ile. Previous work has shown that dipeptides bearing either type 1 or type 2 N-terminal residues act as weak but specific inhibitors of the N-end rule pathway. We took advantage of the two-site architecture of Ubr1p to explore the feasibility of bivalent N-end rule inhibitors, whose expected higher efficacy would result from higher affinity of the cooperative (bivalent) binding to Ubr1p. The inhibitor comprised mixed tetramers of β-galactosidase (βgal) that bore *both* N-terminal Arg (type 1 residue) and N-terminal Leu (type 2 residue) but were resistant to proteolysis. Expression of these constructs in *S. cerevisiae* inhibited the N-end rule pathway much more strongly than the expression of otherwise identical βgal tetramers whose N-terminal residues were exclusively Arg or exclusively Leu. Our findings are the first evidence that the type 1 and type 2 sites of N-recognin are spatially proximal in the 225 kD *S. cerevisiae* Ubr1p. In addition, our results strongly suggest that small bivalent inhibitors of the N-end rule pathway are feasible, and moreover, are expected to be much more potent than their monovalent counterparts. Work to produce such inhibitors is under way. A paper describing these findings is being written up.

**Construction and Use of the Library Expressing Random Fragments of the Green Fluorescent Protein (GFP) and Src Protein Kinase**

The aim of this ongoing *comtoxin* project is to develop a generally applicable method for interfering with the *in vivo* function of a target protein of interest through the expression of a specific fragment of the target that would bind to a conformationally immature (nascent) target protein and thereby either arrest or strongly retard its conformational maturation. As a result, the complex of the target and its self-peptide would be nonfunctional insofar as the target's function is concerned. In addition, the interfering peptide would be equipped with a strong N-terminal degradation signal (the N-degron, previously studied by this laboratory). A permanently or transiently stable complex between a target protein and its self-peptide bearing an N-degron would be targeted for processive degradation by the N-end rule pathway, one of the pathways of the ubiquitin system. Once developed, this method could be applied to effect selective functional inhibition of undesirable proteins such as, for example, overexpressed oncoproteins. This

trans-degradation method would also become a part of the comtoxin approach described in the DOD Idea Grant that supports these studies.

To develop the trans-degradation technique, we chose Green Fluorescent Protein (GFP) as a model target protein. (Our next, clinically relevant, target is Src, a protein tyrosine kinase and an proto-oncoprotein activated in a number of human cancers, including breast cancer.) We carried out the construction of a library of clones which expressed polypeptide fragments derived from random segments of GFP. These fragments were expressed in the yeast *S. cerevisiae* from the $P_{GAL}$ promoter as linear fusions to ubiquitin. As shown previously by this laboratory, the N-terminal ubiquitin moiety of the fusion is cotranslationally cleaved off by ubiquitin-specific proteases, making it possible to expose any desired amino acid residue at the N-terminus of the resulting polypeptide fragment. In developing this approach, we are proceeding in steps, at first constructing a ubiquitin fusion-based library of GFP fragments as such. The second-generation library will have the same fragments bearing the 40-residue N-terminal extension called Arg-$e^{\Delta K}$, which contains the above-described N-degron.

The library expressing random GFP fragments was constructed using DNase I in the presence of $Mn^{2+}$ ions to fragment the GFP cDNA, and specific oligonucleotide adapters to insert the resulting fragments into the expression vector. More recently, we began the construction of analogous libraries with v-Src and c-Src trans-targeting, degron-containing reporter. The v-Src and c-Src cDNAs were obtained from Dr. D. Morgan (Univ. of California, Berkeley). They were tested by sequencing the 5'- and 3'-ends (~200 bp each) of the coding regions, and confirming that the resulting sequences matched those for v-Src and c-Src in the GenBank database. Restriction fragments containing the Src genes were isolated by agarose gel electrophoresis, and are being used to construct random-fragment libraries as described above. Taking advantage of the unique Nco I site at the 5'-end the v-Src cDNA, the v-Src coding region (as a filled-in Nco I-Xho I fragment) was inserted into the integration vector pIP123-CUPR. This placed the Src ORF under the control of the copper-inducible $P_{CUP1}$ promoter. It remains to be determined, in this ongoing project, whether this promoter is sufficiently tight (low enough level of Src in the uninduced state) and sufficiently strong (high enough level of Src in the induced state to cause growth arrest). (In contrast to the GFP-based library, where the screen is being carried out by fluorescence-activated cell sorting (FACS), our readout and the screen with

Src are based on the known toxicity of mammalian Src to the yeast *S. cerevisiae*.) Examination of the three-dimensional structures of Src family tyrosine kinases shows that a loop connecting the SH2 domain to the kinase domain is bound to the SH3 domain of Src and sandwiched between the SH3 domain and the N-terminal half of the kinase when Src is in the inactive state. The kinase-inhibiting interaction requires phosphorylation of Src at its C-terminal region, itself bound to the SH2 domain. This area is a promising candidate for a trans-targeting N-degron.

Our nearest aims in this set of projects are, first, to optimize the library-based interference approach described above, using GFP as both the target and reporter, then to expand this approach to the Src kinase, and to link an N-degron to the interfering fragments identified in both the GFP and Src library screen. This multistep strategy should result in a generally applicable method for metabolically destabilizing (and thereby functionally inhibiting) any protein of interest, including the targets that must be down-regulated for either arresting or selectively killing the breast adenocarcinoma cells.

Enclosed with this Progress Report are our recently published papers (1998-99; one set), and the CV of PI. The two manuscripts-in-preparation mentioned above should be finished in less than 2 months. The Progress Report describes projects that were supported, either partially or entirely, by the DOD Idea Grant during its first year.

March 1999

# CURRICULUM VITAE


**Name:** Alexander J. Varshavsky

**Date and Place of Birth:** November 8, 1946, Moscow, Russia
**Citizenship:** U. S. citizen. Soc. Sec. No. 011-58-8095

**Address:** Division of Biology, 147-75
California Institute of Technology
1200 East California Boulevard
Pasadena, CA 91125

**Telephone:** 626-395-3785 (office); 818-541-9791 (home)
**Fax:** 626-440-9821 (office fax); 818-248-5245 (home fax)
**Email:** avarsh@cco.caltech.edu


**Academic Appointments and Education:**

| | |
|---|---|
| 1970 | B. S. in Chemistry, Department of Chemistry, Moscow University, Moscow, Russia |
| 1973 | Ph. D. in Biochemistry, Institute of Molecular Biology, Moscow, Russia |
| 1973-1976 | Research Fellow, Institute of Molecular Biology, Moscow, Russia |
| 1977-1980 | Assistant Professor of Biology, Department of Biology, M. I. T., Cambridge, Massachusetts |
| 1980-1986 | Associate Professor of Biology, Department of Biology, M. I. T., Cambridge, Massachusetts |
| 1986-1992 | Professor of Biology, Department of Biology, M. I. T., Cambridge, Massachusetts |
| 1992-present | Howard and Gwen Laurie Smits Professor of Cell Biology, Division of Biology, California Institute of Technology, Pasadena, California |

**Other Appointments:**

| | |
|---|---|
| 1983-1987 | Member, Molecular Cytology Study Section, NIH |

**Honors:**

| | |
|---|---|
| 1987 | Elected to the American Academy of Arts and Sciences |
| 1995 | Elected to the U. S. National Academy of Sciences |
| 1998 | Merit Award, National Institutes of Health |
| 1998 | Novartis-Drew Award in Biomedical Science |
| 1999 | Gairdner International Award (shared with Avram Hershko) |

**Publications:** next page

## Selected publications (1968-present)

(grouped by the fields; numbered chronologically;
a paper that encompasses two fields is cited in one of them)

### Chromosome Structure and Gene Expression

1. Varshavsky, A. Regulation of synthesis of genetic repressors in bacteria. **Mol. Biol.** (Russia) 2:13-20 (1968).
5. Varshavsky, A. and Georgiev, G. P. Clustered arrangement of histones F2a1 and F3 in chromosomal deoxyribonucleoproteins. **Biochim. Biophys. Acta** 281:449-674 (1972).
9. Varshavsky, A. and Georgiev, G. P. Redistribution of histones during unfolding of chromosomal DNA. **Mol. Biol. Reports** 1:143-148 (1973).
12. Varshavsky, A. and Ilyin, Y. V. Salt treatment of chromatin induces redistribution of histones. **Biochim. Biophys. Acta** 340:207-217 (1974).
17. Varshavsky, A. and Bakayev, V. V. Nu-bodies and free DNA in chromatin lacking histone H1. **Mol. Biol. Reports** 2:209-217 (1975).
23. Varshavsky, A., Bakayev, V. V., Chumackov, P. M. and Georgiev, G. P. Minichromosome of simian virus 40: presence of histone H1. **Nucl. Acids Res.** 3:2101-2114 (1976).
27. Varshavsky, A. Structural and functional organization of eukaryotic chromosomes. **Biol. Zentralblatt** 95:301-316 (1976).
30. Bakayev, V. V., Bakayeva, T. G. and Varshavsky, A. Nucleosomes and subnucleosomes: heterogeneity and composition. **Cell** 11:619-630 (1977).
31. Varshavsky, A., Nedospasov, S. A., Bakayev, V. V., Bakayeva, T. G. and Georgiev, G. P. Histone-like proteins in the purified *Escherichia coli* deoxyribonucleoprotein. **Nucl. Acids Res.** 4:2725-2745 (1977).
33. Varshavsky, A., Bakayev, V. V., Nedospasov, S. A. and Georgiev, G. P. On the structure of eukaryotic, prokaryotic and viral chromatin. **Cold Spring Harbor Symp. Quant. Biol.** 42:457-472 (1977).
37. Varshavsky, A., Sundin, O. and Bohn, M. SV40 viral minichromosome: preferential exposure of the origin of replication as probed by restriction endonucleases. **Nucl. Acids Res.** 5:3469-3478 (1978).
38. Varshavsky, A., Sundin, O . and Bohn, M. A 400 bp stretch of SV40 viral DNA that includes the origin of replication is exposed in SV40 minichromosomes. **Cell** 16:453-466(1979).
39. Sundin, O. and Varshavsky, A. Staphylococcal nuclease makes a single nonrandom cut in the SV40 viral minichromosome. **J. Mol. Biol.** 132:535-546 (1979).
45. Levinger, L. and Varshavsky, A. Drosophila heat shock proteins are associated with nuclease-resistant, high salt-resistant nuclear structures. **J. Cell Biol.** 90:793-796 (1981).
50. Barsoum, J., Levinger, L. and Varshavsky, A. On the chromatin structure of the amplified, transcriptionally active gene for dihydrofolate reductase in mouse cells. **J. Biol. Chem.** 257:5274-5282 (1982).
51. Levinger, L. and Varshavsky, A. Protein D1 preferentially binds AT-DNA and is a component of *Drosophila melanogaster* nucleosomes containing AT-rich satellite DNA. **Proc. Natl. Acad. Sci. USA** 79:7152-7156 (1982).
53. Swerdlow, P. and Varshavsky, A. Affinity of HMG17 for a nucleosome is not influenced by the presence of ubiquitin-H2A semihistone but depends on DNA fragment size. **Nucl. Acids. Res.** 11:387-401 (1983).
54. Wu, K., Strauss, F. and Varshavsky, A. Nucleosome arrangement in green monkey α-satellite chromatin: superimposition of nonrandom and apparently random patterns. **J. Mol. Biol.** 170:93-117 (1983).
67. Barsoum, J. and Varshavsky, A. Preferential localization of variant nucleosomes near the 5'-end of the mouse dihydrofolate reductase gene. **J. Biol. Chem.** 260:7688-7697 (1985).
73. Solomon, M. J., Strauss, F. and Varshavsky, A. A mammalian HMG protein recognizes a stretch of six AT base pairs in duplex DNA. **Proc. Natl. Acad. Sci. USA** 83:1276-1289 (1986).

80. Peck, L. J., Millstein, L., Eversole-Cire, P., Gottesfeld, J. M. and Varshavsky, A. Transcriptionally inactive oocyte-type 5S RNA genes of *Xenopus laevis* are complexed with TFIIIA *in vitro*. **Mol. Cell. Biol.** 7:3503-3510 (1987).
91. Winter, E. and Varshavsky, A. A DNA-binding protein that recognizes oligo dA-oligo dT tracts. **EMBO J.** 8:1867-1877 (1989).

## DNA Replication, Gene Amplification, Transporters & Drug Resistance

41. Sundin, O. and Varshavsky, A. Terminal stages of SV40 DNA replication proceed via multiply intertwined catenated dimers. **Cell** 21:103-114 (1980).
43. Varshavsky, A. On the possibility of metabolic control of replicon "misfiring": relationship to emergence of malignant phenotypes in mammalian cell lineages. **Proc. Natl. Acad. Sci. USA** 78:3673-3677 (1981).
44. Varshavsky, A. Phorbol ester dramatically increases incidence of methotrexate-resistant cells: possible mechanisms and relevance to tumor promotion. **Cell** 25:561-572 (1981).
46. Sundin, O. and Varshavsky, A. Arrest of segregation leads to accumulation of highly intertwined catenated dimers: dissection of the final stages of SV40 DNA replication. **Cell** 25:659-669 (1981).
58. Varshavsky, A. Do stalled replication forks synthesize a specific alarmone? **J. Theoret. Biol.** 105:707-714, 1983.
55. Barsoum, J. and Varshavsky, A. Mitogenic hormones and tumor promoters greatly increase the incidence of cells bearing amplified dihydrofolate reductase genes. **Proc. Natl. Acad. Sci. USA** 80:5330-5334 (1983).
56. Varshavsky, A. Diadenosine 5', 5'",-$P^1$,$P^4$-tetraphosphate: a pleiotropically acting alarmone? **Cell** 34:711-712 (1983).
59. Snapka, R. and Varshavsky, A. Loss of unstably amplified dihydrofolate reductase genes from mouse cells is accelerated by hydroxyurea. **Proc. Natl. Acad. Sci. USA** 80:7533-7537 (1983).
63. Roninson, I., Abelson, H. T., Housman, D. E., Howell, N. and Varshavsky, A. Amplification of specific DNA sequences correlates with multidrug resistance in Chinese hamster cells. **Nature** 309:626-628 (1984).
71. Ciccarelli, R. B., Solomon, J. J., Varshavsky, A. and Lippard, S. J. *In vivo* effects of cis- and trans-diaminedichloroplatinum (II) on SV40 chromosomes: differential repair, DNA-protein crosslinking, and inhibition of replication. **Biochemistry** 24:7533-7540 (1985).
72. Gros, P., Croop, J., Roninson, I., Varshavsky, A. and Housman, D. E. Isolation and characterization of DNA sequences amplified in multidrug-resistant hamster cells. **Proc. Natl. Acad. Sci. USA** 83:337-341 (1986).
81. Solomon, M. J. and Varshavsky, A. A nuclease-hypersensitive region forms *de novo* after chromosome replication. **Mol. Cell. Biol.** 7:3822-3825 (1987).
93. McGrath, J. P. and Varshavsky, A. The yeast *STE6* gene encodes a homolog of the mammalian multidrug resistance P-glycoprotein. **Nature** 340:400-404 (1989).

## New Methods

16. Bakayev, V. V., Melnickov, A. A., Osicka, V. A. and Varshavsky, A. Isolation and characterization of chromatin subunits. **Nucl. Acids Res.** 2:1401-1419 (1975).
22. Varshavsky, A., Bakayev, V. V. and Georgiev, G. P. Heterogeneity of chromatin subunits and location of histone H1. **Nucl. Acids Res.** 3:477-492 (1976).
42. Levinger, L., Barsoum, J. and Varshavsky, A. Two-dimensional hybridization mapping of nucleosomes. **J. Mol. Biol.** 146:287-304 (1981).
49. Boyce, F., Sundin, O., Barsoum, J. and Varshavsky, A. New way to isolate SV40 viral minichromosomes: use of a thiol-specific reagent. **J. Virol.** 42:292-296 (1982).
64. Strauss, F. and Varshavsky, A. A protein binds to a satellite DNA repeat at three sites which would be brought into proximity by DNA folding in the nucleosome. **Cell** 37:889-901 (1984).

69. Solomon, M. J. and Varshavsky, A. Formaldehyde-mediated DNA-protein crosslinking: a probe for *in vivo* chromatin structures. **Proc. Nat. Acad. Sci. USA** 82:6470-6474 (1985).
74. Swerdlow, P. S., Finley, D. and Varshavsky, A. Enhancement of immunoblot sensitivity by heating of hydrated filters. **Analyt. Biochem.** 156:147-153 (1986).
76. Snapka, R. M., Kwok, K., Bernard, J. A., Harling, O. and Varshavsky, A. Post-separation detection of nucleic acids and proteins by neutron activation. **Proc. Natl. Acad. Sci. USA** 83:9320-9324 (1986).
82. Varshavsky, A. An electrophoretic assay for DNA-binding proteins. **Meth. Enzymol.** 151:551-565 (1987).
83. Bartel, B. and Varshavsky, A. Hypersensitivity to heavy water: a new conditional phenotype. **Cell** 52:935-941 (1988).
86. Solomon, M. J., Larsen, P. L. and Varshavsky, A. Mapping protein-DNA interactions *in vivo* with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. **Cell** 53:937-947 (1988).
112. Dohmen, R. J., Wu, P. P. and Varshavsky, A. Heat-inducible degron: a method for constructing temperature-sensitive mutants. **Science** 263:1273-1276 (1994).
113. Johnsson, N. and Varshavsky, A. Ubiquitin-assisted dissection of protein transport across membranes. **EMBO J.** 13:2686-2698 (1994).
115. Johnsson, N. and Varshavsky, A. Split ubiquitin as a sensor of protein interactions *in vivo*. **Proc. Natl. Acad. Sci. USA** 91:10340-10344 (1994).
123. Lévy, F., Johnsson, N., Rümenapf, T. and Varshavsky, A. Using ubiquitin to follow the metabolic fate of a protein. **Proc. Natl. Acad. Sci. USA** 93:4907-4912 (1996).
127. Johnson, N. and Varshavsky, A. Split ubiquitin: a sensor of protein interactions *in vivo*. In: **The Yeast Two-Hybrid System** (P. L. Bartel and S. Fields, eds.), pp. 316-332, Oxford University Press, N. Y. (1997).
137. Dünnwald, M., Varshavsky, A. and Johnsson, N. Detection of transient *in vivo* interactions between substrate and transporter during protein translocation into the endoplasmic reticulum. **Mol. Biol. Cell** 10:329-344 (1999).

## The Ubiquitin System and Intracellular Proteolysis

40. Levinger, L. and Varshavsky, A. Separation of nucleosomes containing and lacking ubiquitin-H2A semihistone. **Proc. Natl. Acad. Sci. USA** 77:3244-3248 (1980).
48. Levinger, L. and Varshavsky, A. Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the *Drosophila* genome. **Cell** 28:375-386 (1982).
61. Finley, D., Ciechanover, A. and Varshavsky, A. Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. **Cell** 37:43-55 (1984).
62. Ciechanover, A., Finley, D. and Varshavsky, A. Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85. **Cell** 37:57-66 (1984).
66. Özkaynak, E., Finley, D. and Varshavsky, A. The yeast ubiquitin gene: head-to-tail repeats encoding a polyubiquitin precursor protein. **Nature** 312:663-666 (1984).
70. Finley, D. and Varshavsky, A. The ubiquitin system: functions and mechanisms. **Trends Biochem. Sci.** 10:343-346 (1985).
75. Bachmair, A., Finley, D. and Varshavsky, A. *In vivo* half-life of a protein is a function of its N-terminal residue. **Science** 234:179-186 (1986).
77. Özkaynak, E., Finley, D., Solomon, M. J. and Varshavsky, A. The yeast ubiquitin genes: a family of natural gene fusions. **EMBO J.** 6:1429-1440 (1987).
78. Finley, D., Özkaynak, E. and Varshavsky, A. The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation and other stresses. **Cell** 48:1035-1046 (1987).
79. Jentsch, S., McGrath, J. P. and Varshavsky, A. The yeast DNA repair gene *RAD6* encodes a ubiquitin-conjugating enzyme. **Nature** 329:131-134 (1987).

84. Finley, D., Özkaynak, E., Jentsch, S., McGrath, J. P., Bartel, B., Pazin, M., Snapka, R. M. and Varshavsky, A. Molecular genetics of the ubiquitin system. In: **Ubiquitin** (M. Rechsteiner, ed.), pp. 39-75, Plenum Press, N. Y.(1988).
85. Varshavsky, A., Bachmair, A., Finley, D., Wünning, I. and Gonda, D. The N-end rule of selective protein turnover: mechanistic aspects and functional implications. In: **Ubiquitin** (M. Rechsteiner, ed.), pp. 287-324, Plenum Press, N. Y. (1988).
87. Goebl, M. G., Yochem, J., Jentsch, S., McGrath, J. P., Varshavsky, A. and Byers, B. The yeast cell cycle gene *CDC34* encodes a ubiquitin-conjugating enzyme. **Science** 241:1331-1335 (1988).
88. Bachmair, A. and Varshavsky, A. The degradation signal in a short-lived protein. **Cell** 56:1019-1032 (1989).
89. Chau, V., Tobias, J. W., Bachmair, A., Mariott, D., Ecker, D., Gonda, D. K., and Varshavsky, A. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. **Science** 243:1576-1583 (1989).
90. Finley, D., Bartel, B. and Varshavsky, A. The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. **Nature** 338:394-401 (1989).
92. Gonda, D. K., Bachmair, A., Wünning, I., Tobias, J. W., Lane, W. S. and Varshavsky, A. Universality and structure of the N-end rule. **J. Biol. Chem.** 264:16700-16712 (1989).
94. Balzi, E., Choder, M., Chen, W., Varshavsky, A. and Goffeau, A. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. **J. Biol. Chem.** 265:7464-7471 (1990).
95. Hochstrasser, M. and Varshavsky, A. *In vivo* degradation of a transcriptional regulator: the yeast α2 repressor. **Cell** 61:697-708 (1990).
96. Johnson, E. S., Gonda, D. K. and Varshavsky, A. *Cis-trans* recognition and subunit-specific degradation of short-lived proteins. **Nature** 346:287-291 (1990).
97. Bartel, B., Wünning, I. and Varshavsky, A. The recognition component of the N-end rule pathway. **EMBO J.** 9:3179-3189 (1990).
98. Baker, R. T. and Varshavsky, A. Inhibition of the N-end rule pathway in living cells. **Proc. Natl. Acad. Sci. USA** 88:1090-1094 (1991).
99. Varshavsky, A. Naming a targeting signal. **Cell** 64:13-15 (1991).
100. Hochstrasser, M., Ellison, M. J., Chau, V. and Varshavsky, A. The short-lived Matα2 transcriptional regulator is ubiquitinated *in vivo*. **Proc. Nat. Acad. Sci. USA** 88:4606-4610 (1991).
101. Tobias, J. W. and Varshavsky, A. Cloning and functional analysis of the ubiquitin-specific protease gene *UBP1* of *S. cerevisiae*. **J. Biol. Chem.** 266:12021-12028 (1991).
102. Dohmen, R. J., Madura, K., Bartel, B. and Varshavsky, A. The N-end rule is mediated by the Ubc2 (Rad6) ubiquitin-conjugating enzyme. **Proc. Natl. Acad. Sci. USA** 88:7351-7355 (1991).
103. McGrath, J. P., Jentsch, S. and Varshavsky, A. *UBA1*: an essential yeast gene encoding ubiquitin-activating enzyme. **EMBO J.** 10:227-237 (1991).
104. Tobias, J. W., Shrader, T. E., Rocap, G. and Varshavsky, A. The N-end rule in bacteria. **Science** 254:1374-1377 (1991).
105. Johnson, E. S., Bartel, B., Seufert, W. and Varshavsky, A. Ubiquitin as a degradation signal. **EMBO J.** 11:497-505 (1992).
106. Ota, I. and Varshavsky, A. A gene encoding a putative tyrosine phosphatase suppresses lethality of an N-end rule-dependent mutant. **Proc. Natl. Acad. Sci. USA** 89:2355-2359 (1992).
107. Baker, R. T., Tobias, J. W. and Varshavsky, A. Ubiquitin-specific proteases of *S. cerevisiae:* cloning of *UBP2* and *UBP3*, and functional analysis of the *UBP* gene family. **J. Biol. Chem.** 267:23363-23375 (1992).
108. Varshavsky, A. The N-end rule. **Cell** 69:725-735 (1992).

109. Shrader, T. E., Tobias, J. W. and Varshavsky, A. The N-end rule in *Escherichia coli*: cloning and analysis of the leucyl, phenylalanyl-tRNA-protein transferase gene *aat*. **J. Bact**. 175:4364-4374 (1993).
110. Madura, K., Dohmen, R. J. and Varshavsky, A. N-recognin/Ubc2 interactions in the N-end rule pathway. **J. Biol. Chem** 268:12046-12054 (1993).
111. Ota, I. M. and Varshavsky, A. A yeast protein similar to bacterial two-component regulators. **Science** 262:566-569 (1993).
114. Madura, K. and Varshavsky, A. Degradation of Gα by the N-end rule pathway. **Science** 265:1454-1458 (1994).
116. Johnston, J. A., Johnson, E. S., Waller, P. and Varshavsky, A. Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway. **J. Biol. Chem.** 270:8172-8178 (1995).
118. Baker, R. T. and Varshavsky, A. N-terminal amidase: a new enzyme and component of a targeting complex in the N-end rule pathway. **J. Biol. Chem.** 270:12065-12074 (1995).
119. Varshavsky, A. The world of ubiquitin. **Engineering & Science** 58:26-36 (1995).
120. Johnson, E. S., Ma, P. C. M., Ota, I. M. and Varshavsky, A. A proteolytic pathway that recognizes ubiquitin as a degradation signal. **J. Biol. Chem**. 270:17442-17456 (1995).
121. Dohmen, R. J., Stappen, R., McGrath, J. P., Forrová, H., Kolarov, J., Goffeau, A. and Varshavsky, A. An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. **J. Biol. Chem**. 270:18099-18109 (1995).
122. Varshavsky, A. The N-end rule. **Cold Spring Harbor Symp. Quant Biol**. 60:461-478 (1996).
124. Ghislain, M., Dohmen, R. J., Lévy, F., and Varshavsky, A. Cdc48p interacts with Ufd3p, a WD-repeat protein required for ubiquitin-dependent proteolysis in *Saccharomyces cerevisiae*. **EMBO J.** 15:4884-4899 (1996).
125. Varshavsky, A. The N-end rule: functions, mysteries, uses. **Proc. Natl. Acad. Sci. USA** 93:12142-12149 (1996).
126. Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N., Copeland, N. G. and Varshavsky, A. A mouse amidase specific for N-terminal asparagine: the gene, the enzyme, and their function in the N-end rule pathway. **J. Biol. Chem.** 271:28521-28532 (1996).
128. Varshavsky, A. The N-end rule pathway of protein degradation. **Genes Cells** 2:13-29 (1997).
129. Varshavsky, A. The ubiquitin system. **Trends Biochem. Sci**. 22:383-387 (1997).
130. Varshavsky, A., Byrd, C., Davydov, I. V., Dohmen, R. J., Ghislain, M., Gonzalez, M., Grigoryev, S., Johnson, E. S., Johnsson, N., Johnston, J. A., Kwon, Y. T., Lévy, F., Lomovskaya, O., Madura, K., Rümenapf, T., Shrader, T. E., Suzuki, T., Turner, G. and Webster, A. The N-end rule pathway. In **Ubiquitin and the Biology of the Cell** (D. Finley and J.-M. Peters, eds), Plenum Press, NY, pp. 232-278 (1998).
131. Byrd, C. Turner, G. and Varshavsky, A. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. **EMBO J.** 17:269-277 (1998).
132. Davydov, I. V., Patra, D. and Varshavsky, A. The N-end rule pathway in *Xenopus* egg extracts. **Arch. Biochem. Biophys.** 357:317-325 (1998).
133. Kwon, Y. T., Reiss, Y., Fried, V. A., Hershko, A., Yoon, J. K., Gonda, D. K., Sangan, P., Copeland, N. G., Jenkins, N. A. and Varshavsky, A. The mouse and human genes encoding the recognition component of the N-end rule pathway. **Proc. Natl. Acad. Sci. USA** 95:7898-7903 (1998).
134. Ramos, P. C., Höckendorff, J., Johnson, E. S., Varshavsky, A. and Dohmen, R. J. Ump1p is required for proper maturation of the 20S proteasome and becomes its substrate upon completion of the assembly. **Cell** 92:489-499 (1998).
135. Johnston, J. A., Lévy, F. and Varshavsky, A. Analysis of a conditional degradation signal in yeast and mammalian cells. **Eur. J. Biochem.** 259:244-252 (1999).

136. Kwon, Y. T., Kashina, A. S. and Varshavsky, A. Alternative splicing results in differential expression, activity and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. **Mol. Cell. Biol.** 19:182-193 (1999).
137. Dünnwald, M., Varshavsky, A. and Johnsson, N. Detection of transient *in vivo* interactions between substrate and transporter during protein translocation into the endoplasmic reticulum. **Mol. Biol. Cell** 10:329-344.
138. Kwon, Y. T. Lévy, F. and Varshavsky, A. Bivalent inhibitor of the N-end rule pathway. **J. Biol. Chem.** (in press).
139. Turner, G. C. and Varshavsky, A. Detecting and measuring cotranslational protein degradation *in vivo* (submitted for publication).

## Multitarget Compounds

117. Varshavsky, A. Codominance and toxins: a path to drugs of nearly unlimited selectivity. **Proc. Natl. Acad. Sci. USA** 92:3663-3667 (1995).
132. Varshavsky, A. Codominant interference, antieffectors, and multitarget drugs. **Proc. Natl. Acad. Sci. USA** 95:2094-2099 (1998).

## U. S. Patents

1. Varshavsky, A. *Gene amplification assay for detecting tumor promoters.* Issued Apr. 10, 1984.

2. Snapka, R. M., Kwok, K. S., Bernard, J. A., Harling, O. R. and Varshavsky, A. *Indirect labeling method for post-separation detection of chemical compounds.* Issued Dec. 10, 1991.

3. Bachmair, A., Finley, D. and Varshavsky, A. *Methods for generating desired amino-terminal residues in proteins.* Issued Mar. 3, 1992.

4. Varshavsky, A., Johnson, E. S., Gonda, D. K. and Hochstrasser, M. *Methods for trans-destabilization of specific proteins in vivo and DNA molecules useful thereof.* Issued Jun. 16, 1992.

5. Bachmair, A., Finley, D. and Varshavsky, A. *Methods for producing proteins and polypeptides using ubiquitin fusions.* Issued Jul. 21, 1992.

6. Bachmair, A., Finley, D. and Varshavsky, A. *Methods for in vitro cleavage of ubiquitin fusion proteins.* Issued Mar. 23, 1993.

7. Baker, R. T., Tobias, J. W. and Varshavsky, A. *Nucleic acid encoding ubiquitin-specific proteases.* Issued May 18, 1993.

8. Tobias, J. W. and Varshavsky, A. *Ubiquitin-specific protease.* Issued Feb. 21, 1995.

9. Baker, R. T., Tobias, J. T. and Varshavsky, A. *Ubiquitin-specific proteases.* Issued Feb. 27, 1996.

10. Johnsson, N. and Varshavsky, A. *Split-ubiquitin protein sensor.* Issued Apr. 2, 1996.

11. Wu, P., Dohmen, R. J., Johnston, J. and Varshavsky, A. *Heat-inducible N-degron module.* Issued Jul. 23, 1996.

12. Johnston, J. and Varshavsky, A. *Inhibiting degradation of a degron-bearing protein.* Issued Jun. 9, 1998.

13. Baker, R. T. and Varshavsky, A. *Inhibition of protein degradation in living cells with peptides.* Issued June 16, 1998.

14. Kwon, Y. T. and Varshavsky, A. *Nucleic acid encoding mammalian Ubr1.* Issued Jan. 19, 1999.

15. Varshavsky, A. *Codominance-mediated toxins.* Pending patent.

16. Varshavsky, A. *Novel multitarget drugs and therapeutic applications.* Pending patent.

*Eur. J. Biochem.* **259**, 244–252 (1999) © FEBS 1999

# Analysis of a conditional degradation signal in yeast and mammalian cells

**Frédéric Lévy*, Jennifer A. Johnston† and Alexander Varshavsky**

*Division of Biology, California Institute of Technology, Pasadena, CA, USA*

The N-end rule pathway is a ubiquitin-dependent proteolytic system, the targets of which include proteins that bear destabilizing N-terminal residues. The latter are a part of the degradation signal called the N-degron. Arg-DHFR$^{ts}$, an engineered N-end rule substrate, bears N-terminal arginine (a destabilizing residue) and DHFR$^{ts}$ [a temperature-sensitive mouse dihydrofolate reductase (DHFR) moiety]. Previous work has shown that Arg-DHFR$^{ts}$ is long-lived at 23 °C but short-lived at 37 °C in the yeast *Saccharomyces cerevisiae*. In the present work, we extended this analysis, and found that the degradation of Arg-DHFR$^{ts}$ can be nearly completely inhibited *in vivo* by methotrexate (MTX), a low-$M_r$ ligand of DHFR. In *S. cerevisiae*, Arg-DHFR$^{ts}$ is degraded at 37 °C exclusively by the N-end rule pathway, whereas in mouse cells the same protein at the same temperature is also targeted by another proteolytic system, through a degron in the conformationally perturbed DHFR$^{ts}$ moiety. In mouse cells, MTX completely inhibits the degradation of Arg-DHFR$^{ts}$ through its degron within the DHFR$^{ts}$ moiety, but only partially inhibits degradation through the N-degron. When the N-terminus of Arg-DHFR$^{ts}$ was extended with a 42-residue lysine-lacking extension, termed $e^{\Delta K}$, the resulting Arg-$e^{\Delta K}$-DHFR$^{ts}$ was rapidly degraded at both 23 °C and 37 °C. Moreover, the degradation of Arg-$e^{\Delta K}$-DHFR$^{ts}$, in contrast with that of Arg-DHFR$^{ts}$, could not be inhibited by MTX, suggesting that the metabolic stability of Arg-DHFR$^{ts}$ at 23 °C results, at least in part, from steric inaccessibility of its N-terminal arginine. The N-degron of Arg-DHFR$^{ts}$ is the first example of a portable degradation signal the activity of which can be modulated *in vivo* by a cell-penetrating compound. We discuss implications of this advance and the mechanics of targeting by the ubiquitin system.

*Keywords*: degron; methotrexate; N-end rule; proteolysis; ubiquitin.

A number of regulatory circuits, including those that control the cell cycle, cell differentiation and responses to stress, involve metabolically unstable proteins [1–6]. A short *in vivo* half-life of a regulator provides a way to generate its spatial gradient and allows rapid adjustments of its concentration, or subunit composition, through changes in the rate of its synthesis. A protein can also be conditionally unstable, i.e. long-lived or short-lived, depending on the state of the cell. One example of the latter class are cyclins, a family of related proteins the destruction of which at specific stages of the cell cycle regulates cell division and growth [3,7].

Features of proteins that confer metabolic instability are called degradation signals, or degrons [8]. An essential component of one degradation signal, called the N-degron, is a destabilizing N-terminal residue of a protein [9,10]. A set of N-degrons containing different N-terminal residues which are destabilizing in a given cell yields a rule, termed the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. In eukaryotes, the N-end rule pathway is a part of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, differentiation and responses to stress [2,4,11–13]. In many of these settings, Ub acts through routes that involve the degradation of Ub–protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease [14,15].

In eukaryotes, linear Ub–protein fusions are rapidly cleaved by Ub-specific proteases (UBPs) at the Ub–protein junction, making possible the production of otherwise identical proteins bearing different N-terminal residues, a technical advance that led to the discovery of the N-end rule [9,10]. When mouse dihydrofolate reductase (DHFR), a 20-kDa monomeric protein, was expressed as a Ub-Arg-DHFR fusion in the yeast *Saccharomyces cerevisiae* at 30 °C, the resulting Arg-DHFR (produced through deubiquitylation) was long-lived ($t_{0.5} > 4$ h), even though it bore N-terminal Arg, a destabilizing residue [16]. By contrast, a modified protein, Arg-$e^K$-DHFR (produced from Ub-Arg-$e^K$-DHFR), which bore a 42-residue, lysine-containing extension (denoted as $e^K$) between the N-terminal Arg residue and the first (Val) residue of DHFR, was short-lived *in vivo* ($t_{0.5}$ of $\approx$ 10 min), being degraded by the N-end rule pathway [16].

This and other evidence indicated that the N-degron comprises two essential determinants: a destabilizing N-terminal residue and an internal Lys residue (or residues) of a substrate. The Lys residue is the site of attachment of a multi-Ub chain [16–18]. Arg-$e^K$-DHFR contains a complete N-degron because it bears both an N-terminal destabilizing residue (Arg) and, in the $e^K$ extension, at least one targetable Lys residue. In contrast, the N-degron of Arg-DHFR is incomplete, for at least one of the

*Correspondence to* Alexander Varshavsky, Division of Biology, 147–75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125, USA.
Tel.: +1-626-395-3785, Fax: +1-626-440-9821,
E-mail: avarsh@cco.caltech.edu

**Present address*: Ludwig Institute for Cancer Research, Lausanne Branch, Ch. des Boveresses 155, CH-1066 Epalinges, Switzerland.

†*Present address*: Department of Biological Sciences, Stanford University, CA 94305, USA.

*Abbreviations*: βgal, *Escherichia coli* β-galactosidase; DHFR, mouse dihydrofolate reductase; ID, initial decay; MTX, methotrexate; Ub, ubiquitin; UBP, Ubspecific protease; UPR, ubiquitinproteinreference.

*Note*: F. Lévy and J. A. Johnston contributed equally to the present work.

(Received 28 July 1998; revised 12 October 1998; accepted 13 October 1998)

two mutually non-exclusive reasons: the absence of Lys residues accessible to the targeting complex of the N-end rule pathway and/or poor accessibility of the N-terminal Arg in Arg-DHFR to the same targeting complex. Since the 20-kDa mouse DHFR contains 16 Lys residues, this interpretation presumes that, in a folded DHFR molecule, the lysines are ineffective as ubiquitylation sites, because of the lysine's lack of mobility and/or its distance from a destabilizing N-terminal residue [10].

Previous work [19] described a temperature-sensitive (*ts*) allele of DHFR that was used to construct a heat-inducible N-degron, in a Ub-Arg-DHFR$^{ts}$-Ura3p fusion which contained *S. cerevisiae* Ura3p as a C-terminal reporter. The Ub-Arg-DHFR$^{ts}$ moiety of this fusion was identical with Ub-Arg-DHFR above, except for a Pro → Leu alteration at position 66 of the DHFR moiety. Arg-DHFR$^{ts}$-Ura3p was long-lived at 23 °C but short-lived at 37 °C in yeast. Moreover, the Ub-Arg-DHFR$^{ts}$ module of Ub-Arg-DHFR$^{ts}$-Ura3p was shown to be portable, in that linking it to a protein of interest conferred heat-inducible metabolic instability on the entire fusion [19].

In addition to yielding a new method for the construction of *ts* mutants, the heat-inducible N-degron [19] provided an approach to analyzing the mechanism of targeting by the N-end rule pathway. In carrying out this analysis, described below, we have found that the degradation of Arg-DHFR$^{ts}$ can be inhibited *in vivo* by methotrexate (MTX), a low-$M_r$ ligand of DHFR. We also show that, in mouse cells, Arg-DHFR$^{ts}$ is targeted by both the N-end rule pathway and another proteolytic system, whereas in yeast the same test protein is exclusively an N-end rule substrate. The N-degron of Arg-DHFR$^{ts}$ is the first example of a portable degradation signal whose activity can be modulated *in vivo* by a cell-penetrating compound.

## MATERIALS AND METHODS

### Plasmids for expression in *S. cerevisiae*

The plasmid pJJRtd, which expressed Ub-Arg-DHFR$^{ts}$-ha (Fig. 1, construct I), was based on the vector pRS316 [20]. A 1.0-kb *Eco*RI/*Hin*dIII fragment of pPW58 [8] was ligated into pRS316, followed by the cloning of the $P_{CUP1}$ promoter-containing 0.4 kb *Eco*RI/*Sac*I fragment of pPW48 [19] into the *Eco*RI site, a step that yielded pJJRtd. pJJMtd, which expressed Ub-Met-DHFR$^{ts}$-ha, was prepared by ligating the large *AgeI*/*Afl*II fragment of pJJRtd to a small fragment from the *AgeI*/*Afl*II digest of pLGUbM-DHFRha [21].

### Plasmids for expression in mammalian cells

Mouse DHFR and DHFR$^{ts}$ open reading frames bearing the desired restriction sites at their 5′ and 3′ ends were produced using PCR. All DHFR and DHFR$^{ts}$ moieties were flanked, at their C-termini, with the ha epitope [22] (Fig. 1), and were cloned into the pRc/CMV vector (Invitrogen, San Diego, CA, USA).

Constructs II and III (Fig. 1) were produced by PCR amplification of a fragment encoding Met-DHFR$^{ts}$ and Arg-DHFR$^{ts}$, respectively, using construct I as a template. The N-terminal Met of construct II was specified by the PCR primer

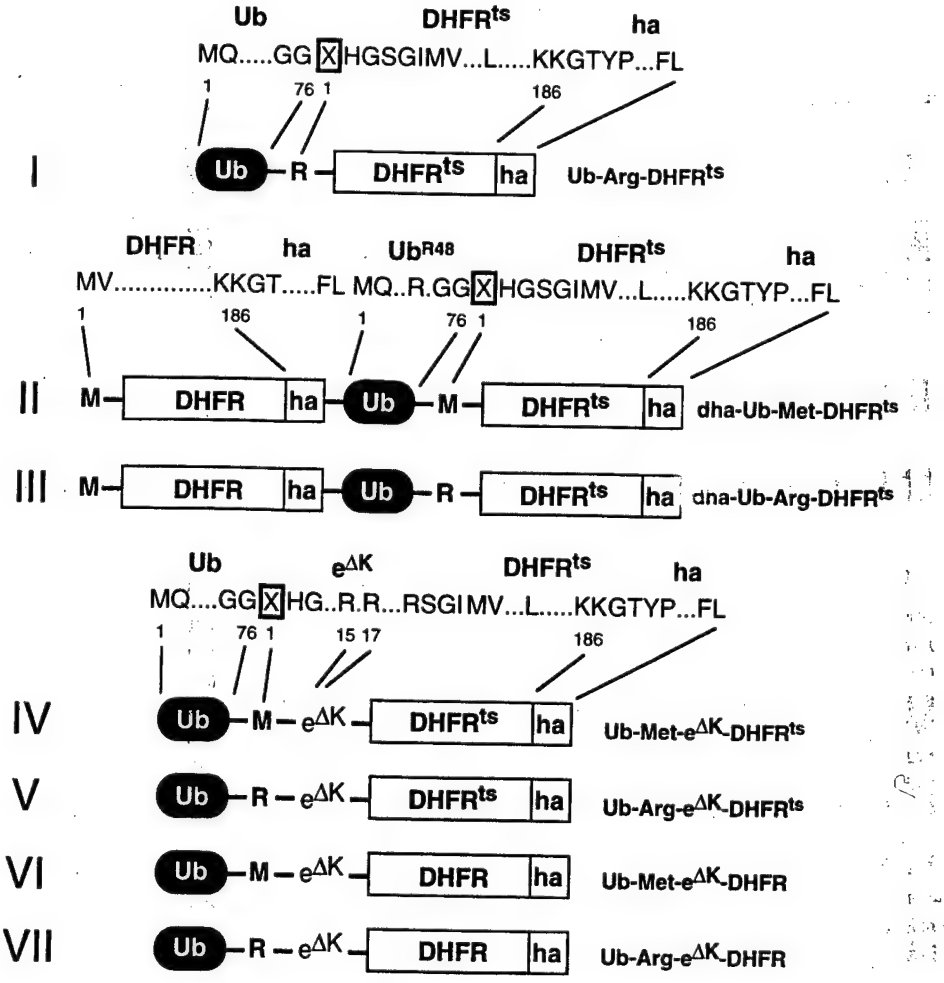



**Fig. 1. Test proteins.** Fusions used in this work contained some of the following moieties: the Ub moiety, either N-terminal (constructs I, IV–VII) or placed between other moieties (constructs II and III); an Arg residue at the junction between Ub and the C-terminal part of the fusion (constructs I, III, V and VII); a Met residue at the same junction (constructs II, IV, and VI); a 42-residue *E. coli* Lac repressor-derived sequence, termed $e^{\Delta K}$ [extension (e) lacking lysines (ΔK)], between Ub and the reporter part of the fusion (see the main text and [22,27]) (constructs IV–VII); DHFRha, a mouse DHFR moiety extended at the C-terminus by a sequence containing the hemagglutinin-derived *ha* epitope (constructs I–VII); DHFR$^{ts}$, a mutant temperature-sensitive DHFR moiety that differs from DHFR by the Pro → Leu alteration at position 66 [19] (see the main text). Amino acid residues are indicated by their single-letter abbreviations. Residues that vary between constructs are boxed in the sequences above the diagrams.

annealing to the 5′ end of the fragment. The PCR products were digested with *Xba*I, blunted with Klenow Pol I, followed by a cut with *Sac*II. A fragment bearing an open reading frame was inserted between the blunt-ended *Not*I site and the *Sac*II site of the plasmid pRc/dhaUbMbgal [23], resulting in the plasmids pRc/dha-Ub-Met-DHFR$^{ts}$ and pRc/dha-Ub-Arg-DHFR$^{ts}$. Met- and Arg-DHFR$^{ts}$ contained the six-residue sequence His-Gly-Ser-Gly-Ile-Met between the N-terminal residue and Val, the first natural residue of DHFR (Fig. 1).

To produce constructs IV–VII (Fig. 1), a fragment encoding Arg-DHFR$^{ts}$ was amplified by PCR, using construct I as template. The PCR product was digested with *Xba*I, and blunted with Klenow Pol I, followed by a *Sac*II cut. The resulting fragment was inserted between the blunt-ended *Not*I site and the *Sac*II site of pRc/UbSβgal, yielding pRc/Ub-Arg-DHFR$^{ts}$. The latter served as a vector for the constructs IV–VII. Constructs IV and V were obtained by amplifying fragments encoding, respectively, Met-e$^{\Delta K}$-DHFR$^{ts}$ and Arg-e$^{\Delta K}$-DHFR$^{ts}$, using the plasmid pLG/Ub$^{V76}$-Val-e$^{\Delta K}$DHFRha [22] as a template. The N-terminal residues Met and Arg were specified by a PCR primer annealing to the 5′ end of the fragment. Digestion of the PCR products with *Sac*II and *Pfl*MI yielded a fragment coding for Met- or Arg-e$^{\Delta K}$ and the first 39 residues of DHFR (DHFR$_{1-39}$). (*Pfl*MI cleaves the DHFR$^{ts}$ open reading frame upstream of the mutated codon at position 66.) This fragment was inserted in-frame between the *Sac*II/*Pfl*MI sites of the plasmid pRc/Ub-Arg-DHFR$^{ts}$, yielding plasmids expressing Ub-Met-e$^{\Delta K}$-DHFR$^{ts}$ and Ub-Arg-e$^{\Delta K}$-DHFR$^{ts}$. Constructs VI and VII were produced by exchanging a *Sna*BI/*Pfl*mI fragment of pRc/dhaUbNe$^{K}$DHFRha with the *Sna*BI/*Pfl*MI fragments derived from constructs IV and V, respectively. This step eliminated the fragment encoding dha-Ub-Asn-e$^{K}$-DHFR$_{1-13}$ of the original plasmid, replacing it, in-frame, with the sequence encoding either Ub-Met-e$^{\Delta K}$-DHFR$_{1-13}$ or Ub-Arg-e$^{\Delta K}$-DHFR$_{1-13}$. (The plasmid pRc/dhaUbNe$^{K}$DHFRha was constructed by replacing the region encoding βgal in pRc/dhaUbXβgal [23] with the region encoding DHFR.)

## Yeast and mouse cell cultures, transfection, and pulse–chase analysis

*S. cerevisiae* were grown and manipulated as described previously [24] and in the legend to Fig. 2. Mouse L cells, a fibroblast-like cell line (ATCC CCL 1.3, American Type

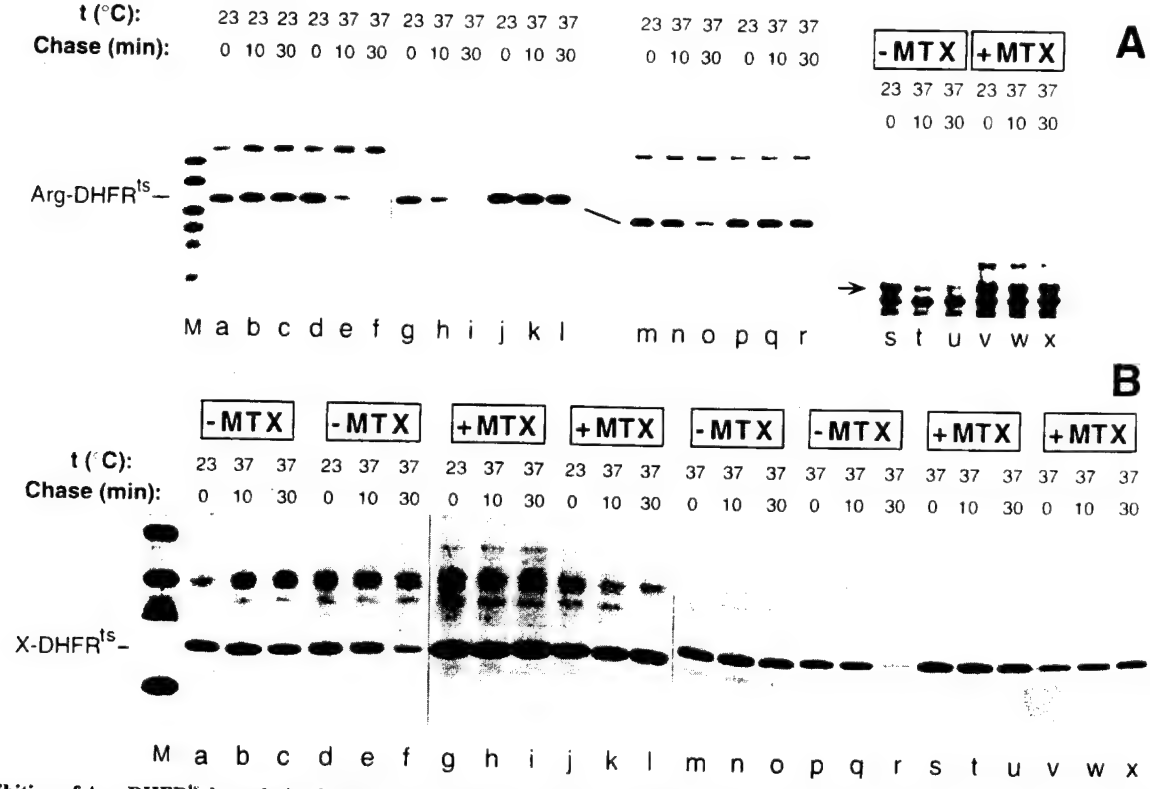



**Fig. 2. Inhibition of Arg-DHFR$^{ts}$ degradation by MTX in S. cerevisiae.** (A) Lane M, molecular mass markers. Lanes a–c, JD47-13c (*UBR1*) cells expressing Arg-DHFR$^{ts}$ (Ub-Arg-DHFR$^{ts}$) were labeled for 4 min at 23 °C with [$^{35}$S]methionine/cysteine, followed by a chase for 10 and 30 min as described [40], then extraction, immunoprecipitation, and Tricine-based SDS/PAGE (12% gel) of Arg-DHFR$^{ts}$ [41]. Lanes d–f, same as lanes a–c, except that immediately after the 4-min pulse cells were shifted to 37 °C, and the chase was carried out at 37 °C. Lanes g–i, same as lanes d–f, but an independent pulse–chase experiment. Lanes j–l, same as lanes g–i, but with JD55 (*ubr1*Δ) *S. cerevisiae*. Lanes m–o, same as lanes g–i, but with KMY613 (*ubc2*Δ) *S. cerevisiae* strain expressing the plasmid-borne *UBC2* gene [42]. Lanes p–r, same as lanes m–o, but with KMY613 (*ubc2*Δ) lacking the plasmid-borne *UBC2*. Lanes s–u, JD47-13c (*UBR1*) cells expressing Arg-DHFR$^{ts}$-ha-Cdc28p (Ub-Arg-DHFR$^{ts}$-ha-Cdc28p) [19] were labeled and processed as described for lanes d–f. Lanes v–x, same as lanes s–u, but with MTX added to the growth medium to a final concentration of 20 μM 30 min before the pulse. (B) Lane M, molecular mass markers. Lanes a–c, JD47-13c (*UBR1*) cells expressing Met-DHFR$^{ts}$ (Ub-Met-DHFR$^{ts}$) were labeled for 4 min at 23 °C with [$^{35}$S]methionine/cysteine. Immediately thereafter, the cells were shifted to 37 °C, followed by a chase at 37 °C for 10 and 30 min, extraction, immunoprecipitation, and SDS/PAGE of Met-DHFR$^{ts}$. Lanes d–f, same as lanes a–c, but with Arg-DHFR$^{ts}$. Lanes g–i, same as lanes a–c, but with MTX added to the growth medium 30 min before the pulse to a final concentration of 20 μM. Lanes j–l, same as lanes g–i, but with Arg-DHFR$^{ts}$. Lanes m–o, same as lanes a–c, except that both the pulse and the chase were at 37 °C. Lanes p–r, same as lanes m–o, but with Arg-DHFR$^{ts}$. Lanes s–u, same as lanes m–o, but with MTX added to the growth medium 30 min before the pulse to a final concentration of 20 μM. Lanes v–x, same as lanes s–u, but with Arg-DHFR$^{ts}$.

Culture Collection, Rockville, MD, USA), were grown as monolayers in Dulbecco's modified Eagle's medium/F12 medium, supplemented with 10% fetal bovine serum, antibiotics, 2 mM L-glutamine and 20 mM Hepes (sodium salt; pH 7.3). The cultures were regularly checked for the absence of mycoplasmas. Transient transfections and pulse–chase analyses were performed as described [23]. For the pulse–chases at 23 °C, cells were transferred to this temperature 30 min before labeling for 10 min with 0.1 mCi (1 Ci = 37 GBq) of [$^{35}$S]EXPRESS (New England Nuclear, Boston, MA, USA). Cells were chased at either 23 °C or 37 °C in the complete medium containing 0.1 mg·mL$^{-1}$ cycloheximide. All media were pre-equilibrated at the desired temperature. In experiments that involved MTX, the drug was added 30 min before labeling and was present at a final concentration of 20 μM (diluted from 20 mM MTX in 0.9 M potassium/sodium phosphate, pH 7.3). The samples were analyzed by SDS/PAGE (15% gel), followed by autoradiography and quantitation using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA, USA).

To distinguish between partial half-lives that approximate the slopes of different regions of a non-exponential decay curve, a generalized half-life term, $t_{0.5}^{y-z}$, was used, where 0.5 denotes the parameter's half-life aspect and $y$–$z$ denotes the relevant time interval, from $y$ to $z$ min of chase. Another term [23], initial decay (ID), expressed as a percentage, equals 1 minus the ratio of the amount of a radiolabeled X-DHFR test protein (normalized against the reference protein, Met-DHFR-Ub; Fig. 1) at the end of the pulse (time 0) to the (normalized) amount of a radiolabeled long-lived X-DHFR such as Met-DHFR. As ID may depend on the duration of a pulse, a superscript, in ID$^y$, invokes the pulse time explicitly [23].

## RESULTS

### Inhibition of Arg-DHFR$^{ts}$ degradation by MTX in *S. cerevisiae*

The major test protein of this work was Arg-DHFR$^{ts}$ (produced from Ub-Arg-DHFR$^{ts}$) (Fig. 1, construct I), identical with the Arg-DHFR$^{ts}$ moiety of a larger fusion described in the introduction. The test proteins bore C-terminal *ha*, a hemagglutinin-derived epitope tag recognized by a monoclonal antibody [22] (Fig. 1; the *ha* epitope is omitted below in the names of specific test proteins). For the analysis in *S. cerevisiae*, Arg-DHFR$^{ts}$ was expressed from the copper-inducible $P_{CUP1}$ promoter in a low-copy vector, and the metabolic fate of Arg-DHFR$^{ts}$ was determined in a pulse–chase assay (see Materials and methods). Similarly to the previously examined Arg-DHFR$^{ts}$-Ura3p and Arg-DHFR$^{ts}$-Cdc28p [19], Arg-DHFR$^{ts}$ was a long-lived protein at 23 °C ($t_{0.5} > 4$ h) but a short-lived protein at 37 °C ($t_{0.5} \approx$ 4 min) (Fig. 2A, lanes a–c vs. d–f). The degradation of Arg-DHFR$^{ts}$ at 37 °C was carried out by the N-end rule pathway, as indicated by (a) the requirement for Ubr1p (E3), the recognition component of this pathway (Fig. 2A, lanes g–i vs. j–l), (b) the requirement for Ubc2p (E2), the relevant Ub-conjugating enzyme (Fig. 2A, lanes m–o vs. p–r), and (c) the long half-life of Met-DHFR$^{ts}$, an otherwise identical protein that bore N-terminal Met, a stabilizing residue, instead of Arg (Fig. 2B, lanes a–c vs. d–f).

Previous work [27] has shown that the degradation of Arg-e$^K$-DHFR by the N-end rule pathway in an extract from rabbit reticulocytes can be inhibited by the folate analog MTX, a high-affinity DHFR ligand ($K_d \approx 10$ pM). It was far from clear whether MTX would have the same effect *in vivo*, in part because the purified N-end rule substrates added to the extract contained the folded DHFR moieties [27]. By contrast, a nascent DHFR-based substrate is unable to bind MTX until after the folding of DHFR [28], but can be targeted by the N-end rule pathway *in vivo* at any time, possibly even during translation [16].

The metabolic fate of Arg-DHFR$^{ts}$ at 37 °C was monitored in *S. cerevisiae* in the absence or presence of 20 μM MTX in the medium. MTX was found to be an efficacious inhibitor of Arg-DHFR$^{ts}$ degradation in *S. cerevisiae* (Fig. 2B, lanes d–f and p–r vs. j–l and v–x). MTX inhibited the degradation of Arg-DHFR$^{ts}$ regardless of whether it was added to the growth medium 15 min or 24 h before the pulse–chase assay (data not shown; the growth of cells was not affected significantly by 20 μM MTX for at least 24 h). The protective effect of MTX on Arg-DHFR$^{ts}$ was essentially the same regardless of whether the pulse labeling (followed by the 37 °C chase) was carried out at 23 °C (Fig. 2B, lanes d–f vs. j–l) or at 37 °C (Fig. 2B, lanes p–r vs. v–x).

The degradation (at 37 °C) of Arg-DHFR$^{ts}$-Cdc28p, which differed from Arg-DHFR$^{ts}$ (produced from Ub-Arg-DHFR$^{ts}$) by the presence of the Cdc28p kinase moiety, was also inhibited by MTX (Fig. 2A, lanes s–u vs. v–x). Arg-DHFR$^{ts}$-Cdc28p, which was long-lived at 23 °C, or at 37 °C in *ubr1Δ* cells, could function as a Cdc28p kinase *in vivo* [19]. As Arg-DHFR$^{ts}$-Cdc28p was short-lived at 37 °C in *UBR1 S. cerevisiae* (Fig. 2A, lanes s–u), the cells that also carried a mutant (*ts*) *cdc28* allele [19] were not viable at 37 °C, owing to the absence of the essential Cdc28p kinase at this temperature. The addition of MTX (at 20 μM) was found to rescue these cells at 37 °C through the inhibition of degradation of Arg-DHFR$^{ts}$-Cdc28p (Fig. 2A, lanes s–u vs. v–x, and data not shown). The effect of MTX was specific for DHFR-containing substrates, as degradation of unrelated test proteins, for example Arg-βgal (Ub-Arg-βgal), by the N-end rule pathway was unimpaired in the presence of MTX (data not shown).

### DHFR as an N-end rule substrate in mammalian cells

DHFR-based substrates of the N-end rule pathway were used in the earlier studies with *S. cerevisiae* [16,19], but have not been examined in mammalian cells. We used L cells, a mouse fibroblast-like cell line, and the ubiquitin/protein/reference (UPR) technique, which increases the accuracy of pulse–chase assays [26] by providing a reference protein [23]. In the UPR technique, Ub is located between a protein of interest and a reference protein in a linear fusion (e.g. construct II in Fig. 1). This fusion is cleaved, cotranslationally or nearly so, by UBPs at the last residue of Ub, producing equimolar amounts of the protein of interest and the reference protein bearing a C-terminal Ub moiety. If both the reference protein and the protein of interest are immunoprecipitated in a pulse–chase assay, the relative amounts of the protein of interest can be normalized against the reference in the same sample. The UPR technique can thus compensate for the scatter of immunoprecipitation yields, sample volumes and other sources of sample-to-sample variation [23].

The UPR constructs of the present work were fusions containing the metabolically stable Met-DHFR-Ub moiety as a reference protein (termed dha-Ub) and a DHFR-based N-end rule substrate such as, for example, Arg-DHFR$^{ts}$ as a protein of interest (Fig. 1, construct III). To preclude the possibility that the C-terminal Ub moiety of dha-Ub could function as a ubiquitylation/degradation signal, the Lys48 residue of Ub (a major site of isopeptide bonds in multi-Ub chains [17,22,29]) was converted into Arg, which cannot be ubiquitylated [12].

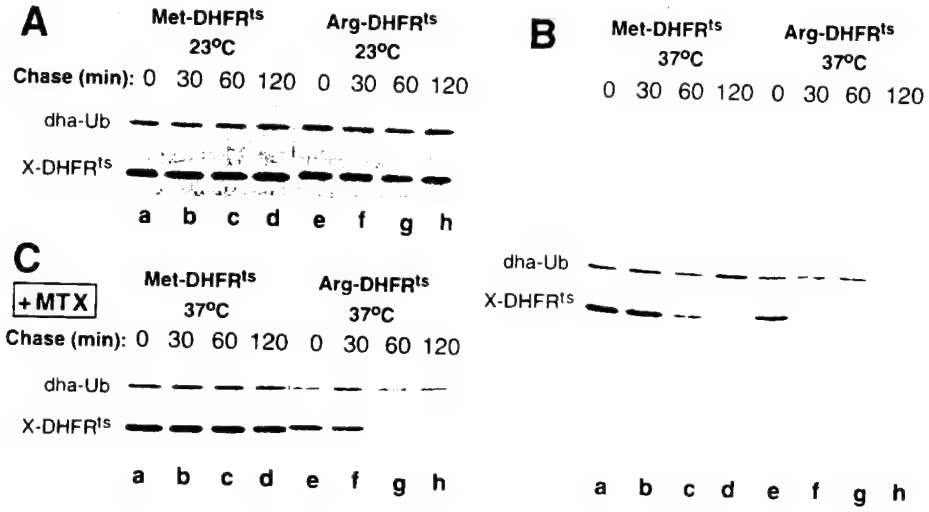



**Fig. 3. Inhibition of Arg-DHFR$^{ts}$ degradation in mouse cells by MTX.** (A) Lanes a–d, mouse L cells transiently expressing a UPR-based version of Met-DHFR$^{ts}$ (Met-DHFR-ha-Ub$^{R48S}$-Met-DHFR$^{ts}$-ha; see Fig. 1, construct II) were labeled at 23 °C for 10 min with [$^{35}$S]methionine/cysteine, followed by a chase, also at 23 °C, for 30, 60 and 120 min in the presence of cycloheximide, extraction, immunoprecipitation, and SDS/PAGE of Met-DHFR$^{ts}$-ha and Met-DHFR-ha-Ub$^{R48S}$. Lanes e–h, same as lanes a–d but with Arg-DHFR$^{ts}$ (Met-DHFR-ha-Ub$^{R48S}$-Arg-DHFR$^{ts}$-ha). (B) Lanes a–d, same as lanes a–d in (A), except that both the pulse and the chase were at 37 °C. Lanes e–h, same as lanes a–d, but with Arg-DHFR$^{ts}$-ha (Met-DHFR-ha-Ub$^{R48S}$-Arg-DHFR$^{ts}$-ha). (C) Lanes a–d, same as lanes a–d in (B) but with 20 μM MTX in the medium. Lanes e–h, same as lanes a–d, but with Arg-DHFR$^{ts}$ (Met-DHFR-ha-Ub$^{R48S}$-Arg-DHFR$^{ts}$-ha). The bands of Arg/Met-DHFR$^{ts}$-ha and Met-DHFR-ha-Ub$^{R48S}$ (denoted as dha-Ub) are indicated. A larger area of the gel is shown for lanes a–h in (B), to illustrate the overall immunoprecipitation pattern.

These fusions were expressed from the cytomegalovirus early promoter, $P_{CMV}$, in mouse L cells. The cells were labeled for 10 min at either 23 °C or 37 °C with [$^{35}$S]methionine/cysteine, followed by a chase of 30, 60 and 120 min at the same temperature in the presence of cycloheximide, immunoprecipitation with anti-ha monoclonal antibody, and analysis/quantitation of immunoprecipitated proteins by SDS/PAGE and PhosphorImager, using UPR (see Materials and methods). Pulse–chase assays in which cycloheximide (a translation inhibitor) was omitted yielded similar results (data not shown).

Both Met-DHFR$^{ts}$ and Arg-DHFR$^{ts}$ were long-lived in mouse cells at 23 °C ($t_{0.5} > 10$ h) (Fig. 3A, lanes a–h, and Fig. 4A). By contrast, Arg-DHFR$^{ts}$ was short-lived at 37 °C in mouse cells: its half-life, determined between 0 and 60 min of chase and denoted as $t_{0.5}^{0-60}$ [23], was ≈ 10 min (Fig. 3B, lanes e–h, and Fig. 4A). In addition, a large fraction (≈ 30%) of the pulse-labeled Arg-DHFR$^{ts}$ (but not of Met-DHFR$^{ts}$) was degraded during the 10-min pulse at 37 °C, as could be seen from comparing the UPR-normalized amounts of Arg-DHFR$^{ts}$ at time 0 (the beginning of the chase) at 23 °C and 37 °C (Fig. 4A). In contrast with a conventional pulse–chase assay, the use of UPR in a setting where a protein can be made either short-lived or long-lived allows the detection and measurement of proteolysis not only during the chase but during the pulse as well [23]. The extent of degradation of a protein during the pulse is denoted as ID$^1$ [23]. In the present context, the variable ID$^1$ is defined as the extent of degradation of a radiolabeled short-lived protein (Arg-DHFR$^{ts}$ at 37 °C) at the end of a pulse relative to the amount of a nearly identical protein (Met-DHFR$^{ts}$ at 37 °C) that is long-lived under the same conditions.

A large fraction, but not all, of the Arg-DHFR$^{ts}$ degradation in mouse cells was carried out by the N-end rule pathway. This could be seen by comparing the UPR-based decay curves of Met-DHFR$^{ts}$ and Arg-DHFR$^{ts}$ at 37 °C (Fig. 3B lanes a–d vs. lanes e–h and Fig. 4A). Specifically, not only Arg-DHFR$^{ts}$ but also Met-DHFR$^{ts}$ were metabolically unstable at 37 °C in mouse cells, the corresponding $t_{0.5}^{0-60}$ being ≈ 10 min and ≈ 48 min, respectively (Fig. 4A). As Met is a stabilizing residue in the N-end rule [23,30], we infer that Arg-DHFR$^{ts}$ was targeted not only by the N-end rule pathway, but also by another proteolytic system, which recognized a structural feature that resulted from a conformational perturbation of the DHFR$^{ts}$ moiety at 37 °C. This aspect of Arg-DHFR$^{ts}$ acted as a degron in mouse cells but not in yeast, inasmuch as the same protein was targeted exclusively by the N-end rule pathway in *S. cerevisiae* at 37 °C (Fig. 2A lanes g–i vs. lanes j–l). The nature of this degradation signal in the conformationally perturbed (at 37 °C) DHFR$^{ts}$ moiety is not known.

## Inhibition of both degradation signals of Arg-DHFR$^{ts}$-ha by MTX in mouse cells

Mouse L cells expressing Met-DHFR$^{ts}$ (dha-Ub-Met-DHFR$^{ts}$) or Arg-DHFR$^{ts}$ (dha-Ub-Arg-DHFR$^{ts}$) were incubated at 37 °C for 30 min with 20 μM MTX in the medium, followed by a 10-min pulse with [$^{35}$S]methionine/cysteine, a chase, and SDS/PAGE analysis. The degradation of Met-DHFR$^{ts}$ at 37 °C, which is mediated by a non-N-degron in the conformationally perturbed DHFR$^{ts}$ moiety (see above), was virtually completely inhibited by MTX (Fig. 3B lanes a–d vs. Fig. 3C lanes a–d and Fig. 4B). Specifically, the $t_{0.5}^{0-60}$ of Met-DHFR$^{ts}$ at 37 °C was ≈ 48 min in the absence of MTX and > 10 h in the presence of MTX (Fig. 4B). The much more rapid degradation of Arg-DHFR$^{ts}$ carried out by both the N-end rule pathway and the other proteolytic pathway was strongly but incompletely inhibited by MTX (Fig. 3B lanes e–h vs. Fig. 3C lanes e–h and Fig. 4B). Specifically, in the absence of MTX, the $t_{0.5}^{0-60}$ of Arg-DHFR$^{ts}$ was ≈ 10 min (ID$^{10}$ of 36%) (Fig. 4B). In the presence of MTX, these variables changed to $t_{0.5}^{0-60}$ ≈ 37 min and ID$^{10}$ of 29% (Fig. 4B).

A comparison of the corresponding decay curves (Fig. 4), made more reliable by the increased accuracy of a UPR-based pulse–chase assay, suggested one reason for the leaky inhibition of the N-degron of Arg-DHFR$^{ts}$ by MTX. Specifically, the ID$^{10}$ of Met-DHFR$^{ts}$ (i.e. the extent of Met-DHFR$^{ts}$ degradation during the pulse) was < 10% in the absence of MTX, whereas

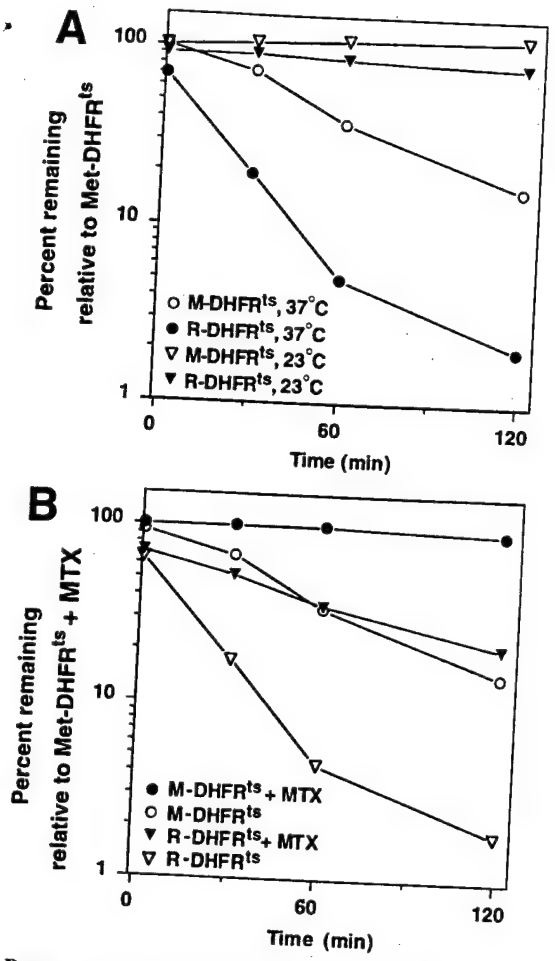



**Fig. 4. Decay curves of Met-DHFR$^{ts}$ and Arg-DHFR$^{ts}$ at 23 °C and 37 °C in mouse cells, in the absence (A) and presence of MTX (B).** These curves were derived from the UPR-based electrophoretic data (Figs 3 and 5 and analogous evidence), as described in Materials and methods.

the ID$^{10}$ of Arg-DHFR$^{ts}$ under the same conditions was ≈ 36%. In other words, more than one-third of labeled Arg-DHFR$^{ts}$ molecules were degraded during the 10-min pulse (Fig. 4B). Moreover, the ID$^{10}$ of Arg-DHFR$^{ts}$ remained high (≈ 30%) even in the presence of MTX (Fig. 4B). This result suggested that a large fraction of the newly formed Arg-DHFR$^{ts}$ molecules was targeted for degradation by the N-end rule pathway before the conformation of Arg-DHFR$^{ts}$ was mature enough to allow high-affinity binding of MTX (see the Discussion).

### An exposed N-terminus renders Arg-DHFR$^{ts}$ constitutively short-lived

To address the relative contributions of the first (N-terminal Arg) and the second (internal Lys) determinant of the N-degron in Arg-DHFR$^{ts}$, we extended its N-terminus with a 42-residue, *Escherichia coli* Lac repressor-derived sequence, termed e$^{\Delta K}$ [extension (e) lacking lysines (ΔK)] [16]. Met-e$^{\Delta K}$-DHFR$^{ts}$ (Ub-Met-e$^{\Delta K}$-DHFR$^{ts}$) (Fig. 1, construct IV), which bore a stabilizing N-terminal residue, was long-lived at 23 °C in mouse cells ($t_{0.5}^{0-60}$ > 10 h) (Fig. 5A, lanes a–c). Arg-e$^{\Delta K}$-DHFR$^{ts}$ (Ub-Arg-e$^{\Delta K}$-DHFR$^{ts}$) (Fig. 1, construct V), which bore a destabilizing N-terminal residue, was short-lived at 23 °C ($t_{0.5}^{0-60}$ ≈ 45 min; Fig. 5A, lanes d–f), in contrast with the otherwise identical Arg-DHFR$^{ts}$ that lacked the e$^{\Delta K}$ extension and was long-lived at 23 °C (Fig. 3A, lanes e–h). Note that Arg-e$^{\Delta K}$-DHFR$^{ts}$ was short-lived at 23 °C, in spite of the fact that the e$^{\Delta K}$ extension did not contribute additional Lys residues to Arg-DHFR$^{ts}$, which was long-lived at 23 °C.

At 37 °C, Met-e$^{\Delta K}$-DHFR$^{ts}$ was approximately as short-lived as Met-DHFR$^{ts}$; the degradation of these proteins was carried out by a non-N-end rule pathway that targeted a degron ($d_\gamma$) in the conformationally perturbed DHFR$^{ts}$ moiety (Fig. 5B, lanes a–c; compare with Fig. 3B, lanes a–d). In contrast, the degradation of Arg-e$^{\Delta K}$-DHFR$^{ts}$ at 37 °C was much faster than that of Met-e$^{\Delta K}$-DHFR$^{ts}$, the difference in rate being due to the targeting of Arg-e$^{\Delta K}$-DHFR$^{ts}$ by both the N-end rule pathway

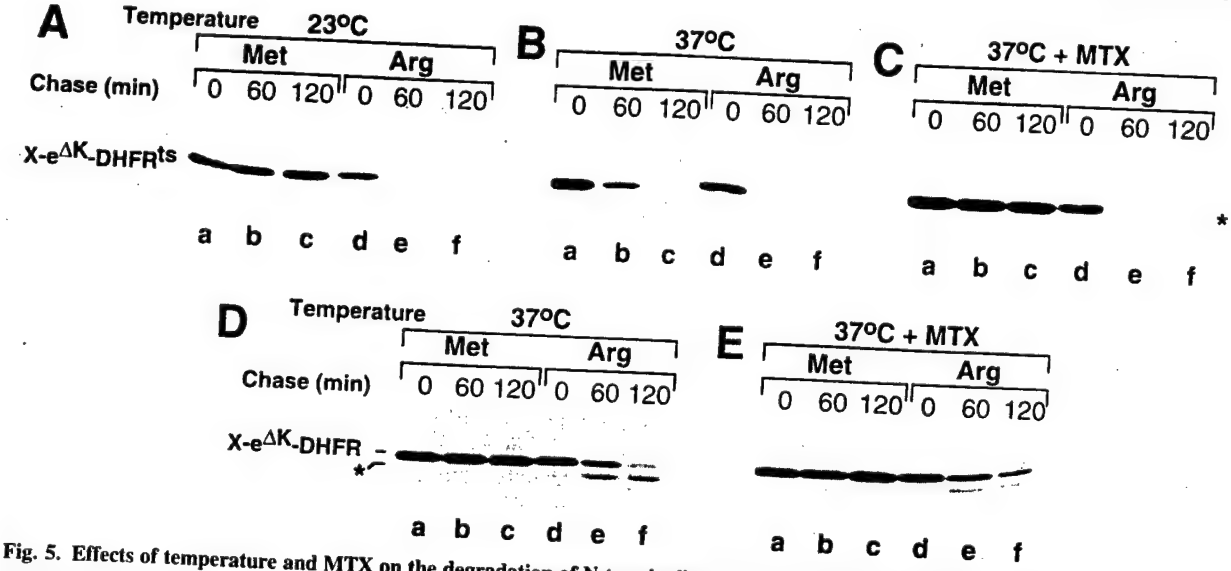



**Fig. 5. Effects of temperature and MTX on the degradation of N-terminally extended DHFR$^{ts}$ and DHFR in mouse cells.** (A) Lanes a–c, mouse L-cells transiently expressing Met-e$^{\Delta K}$-DHFR$^{ts}$ (Fig. 1, construct IV) were labeled at 23 °C for 10 min with [$^{35}$S]methionine/cysteine, followed by a chase, also at 23 °C, for 60 and 120 min in the presence of cycloheximide, extraction, immunoprecipitation, and SDS/PAGE analysis. Lanes d–f, same as lanes a–c but with Arg-e$^{\Delta K}$-DHFR$^{ts}$. (B) Lanes a–f, same as lanes a–f in (A), but the pulse–chase was carried out at 37 °C. (C) Lanes a–c, same as lanes a–c in (B), but the pulse–chase of Met-e$^{\Delta K}$-DHFR$^{ts}$ was carried out in the presence of 20 μM MTX in the growth medium. Lanes d–f, same as lanes a–c, but with Arg-e$^{\Delta K}$-DHFR$^{ts}$. (D) Lanes a–c, same as lanes a–c in (B), but with Met-e$^{\Delta K}$-DHFR (the wild-type DHFR moiety). Lanes d–f, same as lanes a–c, but with Arg-e$^{\Delta K}$-DHFR. The asterisk indicates a cleavage product derived from a minority of short-lived Arg-e$^{\Delta K}$-DHFR$^{ts}$ and Arg-e$^{\Delta K}$-DHFR molecules during their targeting by the N-end rule pathway (see the main text). (E) Lanes a–c, same as lane a–c in (D), but in the presence of 20 μM MTX. Lanes d–f, same as lanes a–c but with Arg-e$^{\Delta K}$-DHFR.

and the $d_s$-recognizing pathway (Fig. 5B, lanes d–f). A minor proteolytic fragment of Arg-$e^{\Delta K}$-DHFR$^{ts}$, visible in Fig. 5B–E (lanes d–f), was specific to N-end rule substrates bearing an $e^{\Delta K}$-type extension; an analogous fragment was also observed with X-$e^{K}$-βgal-based N-end rule substrates [16].

We also examined the metabolic fate of Met-$e^{\Delta K}$-DHFR (Ub-Met-$e^{\Delta K}$-DHFR) and Arg-$e^{\Delta K}$-DHFR (Ub-Arg-$e^{\Delta K}$-DHFR), which were identical with the substrates above except that they bore the wild-type mouse DHFR moiety (Fig. 1, constructs VI and VII). Met-$e^{\Delta K}$-DHFR was long-lived at both 23 °C and 37 °C in mouse cells (Fig. 5D, lanes a–c, and data not shown), indicating that the heat-activated $d_s$-type degron of substrates containing the DHFR$^{ts}$ moiety (e.g. Fig. 4A) resulted from the Pro66 gr; Leu66 mutation that yielded this moiety. The degradation of Arg-$e^{\Delta K}$-DHFR was mediated exclusively by the N-end rule pathway (Fig. 5D, lanes d–f vs. lanes a–c), unlike the double-pathway degradation of Arg-$e^{\Delta K}$-DHFR$^{ts}$ at 37 °C (Fig. 5C, lanes d–f). In contrast with the results with Arg-DHFR$^{ts}$ and Met-DHFR$^{ts}$, the presence of MTX did not result in a significant stabilization of Arg-$e^{\Delta K}$-DHFR (Fig. 5D lanes d–f vs. Figure 5E lanes d–f).

## DISCUSSION

We report the following results.

(a) Arg-DHFR$^{ts}$ is long-lived at 23 °C but short-lived at 37 °C in both *S. cerevisiae* and mouse cells.

(b) In yeast, Arg-DHFR$^{ts}$ is degraded (at 37 °C) exclusively by the N-end rule pathway, whereas in mouse cells the same protein at the same temperature is degraded by another proteolytic pathway as well. The corresponding mouse-specific degron is active at 37 °C but inactive at 23 °C. This degradation signal is a currently unknown feature of the perturbed conformation of the DHFR$^{ts}$ moiety at 37 °C, which differs from the wild-type DHFR moiety by a Pro → Leu alteration at position 66.

(c) MTX, a low-$M_r$ ligand of DHFR ($K_d \approx 10$ pM), inhibits the degradation of Arg-DHFR$^{ts}$ (at 37 °C) nearly completely in yeast and partially in mouse cells.

(d) The ability of MTX to inhibit the degradation of Arg-DHFR$^{ts}$ *in vivo* is retained when Arg-DHFR$^{ts}$ is utilized as a portable degron that confers metabolic instability on linked unrelated proteins in yeast.

(e) Using the UPR technique [23] to determine the metabolic fate of Arg-DHFR$^{ts}$ in mouse cells, we found that ≈ 36% of the labeled Arg-DHFR$^{ts}$ molecules are degraded during the 10-min pulse, and that the presence of MTX decreases this fraction only slightly, to ≈ 30%. This finding accounts for the leaky inhibition of Arg-DHFR$^{ts}$ degradation by MTX in mouse cells. Specifically, this finding suggests that a large fraction of the nascent Arg-DHFR$^{ts}$ is targeted, in mouse cells, by the N-end rule pathway before the conformation of the DHFR$^{ts}$ moiety is mature enough [28] to allow high-affinity binding of MTX. That the MTX-mediated inhibition of degradation of Arg-DHFR$^{ts}$ in yeast is much less leaky than in mouse cells is but one difference in the detailed properties of the yeast and mammalian N-end rule pathways [see also item (f)].

(f) When the N-terminus of Arg-DHFR$^{ts}$ was extended with a 42-residue lysine-lacking extension, termed $e^{\Delta K}$, the N-degron of the resulting Arg-$e^{\Delta K}$-DHFR$^{ts}$ was active at both 23 °C and 37 °C, unlike the mouse-specific degron, which remained inactive at 23 °C. Moreover, the degradation of Arg-$e^{\Delta K}$-DHFR$^{ts}$, in contrast with that of Arg-DHFR$^{ts}$, could not be inhibited by MTX, suggesting that the inactivity of the N-degron in Arg-DHFR$^{ts}$ at 23 °C results at least in part from inaccessibility of its N-terminal arginine to the targeting complex of the N-end rule pathway. Arg-$e^{\Delta K}$-DHFR, which bore the wild-type DHFR moiety, was also short-lived in mouse cells. Previously, the same protein was found to be long-lived in yeast, unless the $e^{\Delta K}$ extension was replaced by an otherwise identical extension, termed $e^{K}$, which contained Lys residues that functioned as the second determinant of the N-degron [16].

Potential applications of a heat-inducible, MTX-suppressible portable N-degron include its use to produce conditional mutants in homeothermic animals such as mouse. One difficulty in the current approaches to this problem [31–33] is that either a deletion or transcriptional repression of a gene of interest leaves the previously produced gene product unperturbed. If this protein is long-lived and resides in a non-dividing cell, there may be a considerable phenotypic lag between the conditionally introduced genetic change and the actual inactivation or disappearance of a corresponding protein. A degron that can be regulated by a cell-penetrating ligand can be employed to address this problem. Since chronic administration of MTX is toxic to mammals, it will be necessary to construct an analogous DHFR-based N-degron that uses *E. coli* DHFR and is inhibited by trimethoprim, which binds tightly to the *E. coli* but not the mouse DHFR [34]. If a portable N-degron of this kind could be constructed, combining it with a conditional repression of a gene of interest should yield better methods for producing conditional mutants in homeothermic animals.

MTX, through its binding to the substrate pocket of DHFR, stabilizes DHFR conformation [35,36]. This effect was employed in several studies with DHFR as a reporter protein. For example, the binding of MTX to DHFR can preclude, under certain conditions, the translocation of a DHFR-containing fusion protein across biological membranes [37,38]. MTX can partially protect DHFR against *in vitro* proteolysis by thermolysin [38,39]. MTX has also been shown to block the degradation of a targeted multiubiquitylated Arg-$e^{K}$-DHFR in reticulocyte extract [27]. In the latter study, the N-end rule pathway in reticulocyte extract was presented with a prefolded wild-type DHFR moiety. By contrast, in the present work with intact cells, the same pathway could target a DHFR-based N-end rule substrate immediately after (and possibly even during) its synthesis. Further analysis is required to clarify the apparent discrepancy between the previously observed inhibition of degradation of Arg-$e^{K}$-DHFR by MTX *in vitro* (in reticulocyte extract) [27] and the absence of a comparable effect of MTX *in vivo* on this class of substrates, in which N-terminal Arg is constitutively exposed. One difference between the two settings is the presentation of prefolded DHFR moieties in reticulocyte extract compared with a kinetic competition between the folding of newly formed DHFR moieties and their targeting for degradation by the N-end rule pathway.

The use of the UPR technique not only increased the overall accuracy of pulse–chase assays with mouse cells but also showed that ≈ 30% of the newly made Arg-DHFR$^{ts}$ was degraded during the pulse. The finding that the fraction of Arg-DHFR$^{ts}$ degraded during the pulse decreased only slightly in the presence of MTX strongly suggests that most of the labeled Arg-DHFR$^{ts}$ is conformationally immature shortly after synthesis, and therefore cannot form a high-affinity complex with MTX. The Pro → Leu alteration at position 66 that yielded DHFR$^{ts}$ occurred in the region of DHFR that interacts with the aromatic ring of MTX [35], and is likely to have resulted in a decreased affinity of DHFR for MTX. Nonetheless, the test proteins containing the DHFR$^{ts}$ moiety were efficiently retained on an MTX affinity column (data not shown), and their *in vivo* degradation could be specifically inhibited by MTX.

A model of targeting by the N-end rule pathway that accounts for our findings with the heat-inducible N-degron assumes that the main cause of metabolic stability of Arg-DHFR$^{ts}$ at 23 °C is not the absence of a sterically accessible Lys residue within the folded DHFR globule, but rather a steric inaccessibility of the N-terminal Arg residue. In this model, the activation of the previously cryptic N-degron of Arg-DHFR$^{ts}$ at 37 °C is caused by an increased exposure and flexibility of a region bearing the N-terminal Arg, a transition that leads to the binding of Arg by the targeting complex of the N-end rule pathway. According to this interpretation, MTX stabilizes Arg-DHFR$^{ts}$ against degradation at 37 °C not by precluding the conformational mobilization of internal Lys residues, but by precluding a temperature-mediated increase in the exposure of the N-terminal Arg. In addition to being consistent with the available evidence, this model accounts for the otherwise puzzling result that Arg-e$^{\Delta K}$-DHFR$^{ts}$-ha (which differs from Arg-DHFR$^{ts}$ by the presence of e$^{\Delta K}$) is short-lived at both 23 °C and 37 °C, whereas Arg-DHFR$^{ts}$ is short-lived only at 37 °C. Indeed, the e$^{\Delta K}$ extension would result in a temperature-independent enhanced exposure of N-terminal Arg, a change that would be, in this model, sufficient for the temperature-independent activity of the N-degron in Arg-e$^{\Delta K}$-DHFR$^{ts}$. The earlier model proposed by Dohmen *et al.* [19], which presumed that the inactivity of the N-degron in Arg-DHFR$^{ts}$ at 23 °C was caused by the absence of sterically accessible Lys residues, cannot account for the above result.

How can we verify the Arg-exposure model? One way is suggested by the existence, in both yeast and mammals, of the enzymes N-terminal amidase (Nt-amidase) and Arg-tRNA-protein transferase (R-transferase), which chemically modify specific (sterically exposed) N-terminal residues of proteins in the cytosol. In *S. cerevisiae*, the *NTA1*-encoded Nt-amidase deamidates N-terminal Asn or Gln; the *ATE1*-encoded R-transferase conjugates Arg to N-terminal Asp or Glu (reviewed in [10]). A test of the model would involve the expression of, for example, Asn-DHFR$^{ts}$ (identical to Arg-DHFR$^{ts}$ except for the presence of N-terminal Asn) in yeast cells that lack Ubr1p, a component of the N-end rule pathway that recognizes primary destabilizing N-terminal residues such as Arg. (The absence of Ubr1p would preclude the degradation of test proteins.) The model predicts that at 23 °C the N-terminal Asn of Asn-DHFR$^{ts}$ would be deamidated inefficiently or not at all, in contrast with what happens at 37 °C. This can be verified by isolating Asn-DHFR$^{ts}$ from cells incubated at either 23 °C or 37 °C, and determining its N-terminal residue. If the model is correct, the N-terminal sequence must begin largely or entirely with Asn at 23 °C, but would become Arg-Asp-... at 37 °C. (In the latter case, the N-terminal Asp, produced from Asn, would be arginylated by R-transferase.) Experimental verification of this model is under way.

## ACKNOWLEDGEMENTS

We thank Michel Ghislain, Ailsa Webster and Martin Gonzalez for helpful discussions and advice, and Lawrence Peck for comments on the manuscript. This work was supported by the grants to A.V. from the National Institutes of Health (DK39520 and GM31530). F.L. and J.A.J. were supported by fellowships from the Swiss National Fund for Research and the American Cancer Society, respectively.

## REFERENCES

1. Haas, A.J. & Siepman, T.J. (1997). Pathways of ubiquitin conjugation. *FASEB J.* **11**, 1257–1268.
2. Hershko, A. (1997). Roles of ubiquitin-mediated proteolysis in cell cycle control. *Curr. Op. Cell. Biol.* **9**, 788–799.
3. King, R.W., Deshaies, R.J., Peters, J.M. & Kirschner, M.W. (1996). How proteolysis drives the cell cycle. *Science* **274**, 1652–1659.
4. Varshavsky, A. (1997). The ubiquitin system. *Trends Biochem. Sci.* **22**, 383–387.
5. Wilkinson, K.D. (1997). Regulation of ubiquitin-dependent processes by deubiquitinating enzymes. *FASEB J.* **11**, 1245–1256.
6. Pickart, C.M. (1997). Targeting of substrates to the 26S proteasome. *FASEB J.* **11**, 1055–1066.
7. Murray, A. & Hunt, T. (1993) *The Cell Cycle*. Freeman, New York.
8. Varshavsky, A. (1991). Naming a targeting signal. *Cell* **64**, 13–15.
9. Bachmair, A., Finley, D. & Varshavsky, A. (1986). *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186.
10. Varshavsky, A. (1997). The N-end rule pathway of protein degradation. *Genes to Cells* **2**, 13–28.
11. Hicke, L. (1997). Ubiquitin-dependent internalization and down-regulation of plasma membrane proteins. *FASEB J.* **11**, 1215–1226.
12. Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405–439.
13. Scheffner, M., Smith, S. & Jentsch, S. (1998) In *Ubiquitin and the Biology of the Cell* (Peters, J.-M., Harris, J.R. & Finley, D., eds), pp. 65–98. Plenum Press, New York.
14. Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. (1998). The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92**, 367–380.
15. Hilt, W. & Wolf, D.H. (1996). Proteasomes: destruction as a programme. *Trends. Biochem. Sci.* **21**, 96–102.
16. Bachmair, A. & Varshavsky, A. (1989). The degradation signal in a short-lived protein. *Cell* **56**, 1019–1032.
17. Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K. & Varshavsky, A. (1989). A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **243**, 1576–1583.
18. Hill, C.P., Johnston, N.L. & Cohen, R.E. (1993). Crystal structure of a ubiquitin-dependent degradation substrate: a three-disulfide form of lysozyme. *Proc. Natl. Acad. Sci. USA* **90**, 4136–4140.
19. Dohmen, R.J., Wu, P. & Varshavsky, A. (1994). Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science* **263**, 1273–1276.
20. Sikorski, R.S. & Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *S. Cerevisiae*. *Genetics* **122**, 19–27.
21. Johnson, E.S., Bartel, B.W. & Varshavsky, A. (1992). Ubiquitin as a degradation signal. *EMBO J.* **11**, 497–505.
22. Johnson, E.S., Ma, P.C.M., Ota, I.M. & Varshavsky, A. (1995). A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.* **270**, 17442–17456.
23. Lévy, F., Johnsson, N., Rümenapf, T. & Varshavsky, A. (1996). Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl. Acad. Sci. USA* **93**, 4907–4912.
24. Ghislain, M., Dohmen, R.J., Lévy, F. & Varshavsky, A. (1996). Cdc48p interacts with Ufd3p, a WD-repeat protein required for ubiquitin-dependent proteolysis in *Saccharomyces Cerevisiae*. *EMBO J.* **15**, 4884–4899.
25. Baker, R.T. & Varshavsky, A. (1991). Inhibition of the N-end rule pathway in living cells. *Proc. Natl. Acad. Sci. USA* **87**, 2374–2378.
26. Mosteller, R.D. & Goldstein, B.E. (1984). A mathematical model that applies to protein degradation and posttranslational processing of proteins and to analogous processes for other molecules in non-growing and exponentially growing cells. *J. Theor. Biol.* **108**, 597–621.
27. Johnston, J.A., Johnson, E.S., Waller, P.R.H. & Varshavsky, A. (1995). Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway. *J. Biol. Chem.* **270**, 8172–8178.
28. Frieden, C. (1990). Refolding of Escherichia coli dihydrofolate reductase: sequential formation of substrate-binding sites. *Proc. Natl. Acad. Sci. USA* **87**, 4413–4416.
29. Spence, J., Sadis, S., Haas, A.L. & Finley, D. (1995). A ubiquitin mutant

with specific defects in DNA repair and multiubiquitination. *Mol. Cell. Biol.* **15**, 1265–1273.

30. Gonda, D.K., Bachmair, A., Wünning, I., Tobias, J.W., Lane, W.S. & Varshavsky, A. (1989). Universality and structure of the N-end rule. *J. Biol. Chem.* **264**, 16700–16712.
31. Gu, H., Marth, J.D., Orban, P.C., Mossmann, H. & Rajewski, K. (1994). Deletion of a DNA polymerase β gene segment in T cells using cell type-specific gene targeting. *Science* **265**, 103–106.
32. No, D., Yao, T.-P. & Evans, R.M. (1996). Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. USA* **93**, 3346–3351.
33. St-Onge, L., Furth, P.A. & Gruss, P. (1996). Temporal control of the Cre recombinase in transgenic mice by a tetracycline-responsive promoter. *Nucl. Acids Res.* **24**, 3975–3877.
34. Matthews, D.A., Bolin, J.T., Burridge, J.M., Filman, D.J., Volz, K.W. & Kraut, J. (1985). Dihydrofolate reductase: the stereochemistry of inhibitor selectivity. *J. Biol. Chem.* **260**, 392–399.
35. Oefner, C., D'Arcy, A. & Winkler, F.K. (1988). Crystal structure of human dihydrofolate reductase complexed with folate. *Eur. J. Biochem.* **174**, 377–385.
36. Thillet, J., Absil, J., Stone, S.R. & Pictet, R. (1988). Site-directed mutagenesis of mouse dihydrofolate reductase: mutants with increased resistance to methotrexate and trimethoprim. *J. Biol. Chem.* **263**, 12500–12508.
37. Arkowitz, R.A., Joly, J.C. & Wickner, W. (1992). Translocation can drive the unfolding of a preprotein domain. *EMBO J.* **12**, 243–253.
38. Eilers, M. & Schatz, G. (1986). Binding of a specific ligand inhibits import of a purified precursor into mitochondria. *Nature* **322**, 228–232.
39. Klingenberg, O. & Olsnes, S. (1996). Ability of methotrexate to inhibit translocation to the cytosol of dihydrofolate reductase fused to diphtheria toxin. *Biochem. J.* **313**, 647–653.
40. Madura, K. & Varshavsky, A. (1994). Degradation of Gα by the N-end rule pathway. *Science* **265**, 1454–1458.
41. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A. & Seidman, J.G. & Struhl, K. (1996) *Current Protocols in Molecular Biology*. Wiley-Interscience, New York.
42. Madura, K., Dohmen, R.J. & Varshavsky, A. (1993). N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.* **268**, 12046–12054.

*Proc. Natl. Acad. Sci. USA*
Vol. 95, pp. 2094–2099, March 1998
Biochemistry

# Codominant interference, antieffectors, and multitarget drugs

(pharmacology/cancer/codominance/selectivity)

ALEXANDER VARSHAVSKY†

Division of Biology, California Institute of Technology, Pasadena, CA 91125

*Contributed by Alexander Varshavsky, December 19, 1997*

**ABSTRACT** **The insufficient selectivity of drugs is a bane of present-day therapies. This problem is significant for antibacterial drugs, difficult for antivirals, and utterly unsolved for anticancer drugs, which remain ineffective against major cancers, and in addition cause severe side effects. The problem may be solved if a therapeutic agent could have a multitarget, combinatorial selectivity, killing, or otherwise modifying, a cell if and only if it contains a predetermined set of molecular targets and lacks another predetermined set of targets. An earlier design of multitarget drugs [Varshavsky, A. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3663–3667] was confined to macromolecular reagents such as proteins, with the attendant difficulties of intracellular delivery and immunogenicity. I now propose a solution to the problem of drug selectivity that is applicable to small (≤1 kDa) drugs. Two ideas, codominant interference and antieffectors, should allow a therapeutic regimen to possess combinatorial selectivity, in which the number of positively and negatively sensed macromolecular targets can be two, three, or more. The nature of the effector and interference moieties in a multitarget drug determines its use: selective killing of cancer cells or, for example, the inhibition of a neurotransmitter-inactivating enzyme in a specific subset of the enzyme-containing cells. The *in vivo* effects of such drugs would be analogous to the outcomes of the Boolean operations "and," "or," and combinations thereof. I discuss the logic and applications of the antieffector and interference/codominance concepts, and the attendant problem of pharmacokinetics.**

The many successes of pharmacology (1, 2) do not include the problem of cancer. Major human cancers are incurable once they have metastasized. A few relatively rare cancers, such as testicular carcinoma in men, Wilms' kidney tumor, and some leukemias in children, can often be cured through chemotherapy but require cytotoxic treatments of a kind that cause severe side effects and are themselves carcinogenic (2).

The main reason for the failure of cytotoxic therapies is their insufficient selectivity for tumors. For example, treatments with radiation or alkylating agents perturb many functions that are common to all cells. The more selective cytotoxic drugs, for instance, methotrexate, taxol, and etoposide, perturb the functions of specific macromolecular targets (dihydrofolate reductase, microtubules, and topoisomerase II), but these targets are present in both normal and malignant cells (1, 2). Hence the low therapeutic index of anticancer drugs and their systemic toxicity at clinically relevant doses. Because the mitotic activity of cells in a tumor is often lower than the mitotic activity of normal cells in self-renewing tissues such as the bone marrow (3), one might not have expected these drugs to work at all—to have any preference for the killing of cancer cells. That such preference actually exists stems in part from the fact that tumor cells are often perturbed by their mutations into stress-hypersensitive states. Consequently, these tumor cells die an apoptotic death at the level of a drug-imposed metabolic stress that induces apoptosis in some but not in most of the organism's normal cells (3).

With some cancers, cytotoxic therapies are ineffective from the beginning. In other cases, these therapies yield a partial, sometimes clinically complete, but almost invariably transient remission of a cancer, in part because these treatments select for tumor cell variants that retain tumorigenicity but are more resistant to either apoptosis *per se* or a drug that induces apoptosis. Because a significant increase in the drug or radiation dosage is precluded by their low therapeutic index, these therapies become ineffective when resistant clones of malignant cells, selected by a drug treatment, present themselves as a cancer recurrence.

The failure of small cytotoxic drugs to produce a cure for cancer has given rise to other strategies, in particular the insightful suggestion that solid tumors can be targeted by selectively inhibiting neovascularization, a process that these tumors depend on for growing to a clinically significant size (4). Another approach, immunotoxins, involves the linking of a toxin to a ligand such as an antibody or a growth factor that binds to a target on the surface of tumor cells (5). Among the limitations of present-day immunotoxins is their incapacity, on entering a cell, to adjust their toxicity in response to the intracellular protein composition. Yet another approach is to enhance the ability of the immune system to identify and selectively destroy tumor cells. The current revival of this strategy holds the promise of a rational and curative treatment (6). Given the complicated regimens and unsolved problems of immunotherapies, it is clear that this and other recent approaches (7, 8) are motivated in part by the perception that small-drug pharmacology, so successful against bacterial infections, is unlikely to prove effective against cancer. In contrast to this view, the premise of the strategy described in the present work is that small anticancer drugs may become curative and free of severe side effects if a way is found to confer on these compounds a multitarget, combinatorial selectivity.

Most cancers are monoclonal: cell lineages of both the primary tumor and the metastases originate from a single founder cell. This cell is a breakthrough descendant of a cell lineage that has been accumulating mutations for some time, often in proximity to other neoplastic but still nonmetastatic cell lineages within an indolent proliferative lesion such as a benign tumor (9, 10). Given the monoclonality of a cancer, cells of both the primary tumor and the metastases share the initial mutations that yielded the founder cell, even if these cells differ at other loci that accumulated mutations in the course of the later tumor progression. Some of the early mutations are in genes that encode tumor suppressors (9, 11, 12). In most cancers, both alleles of a tumor suppressor gene are inactivated, sometimes through deletions that encompass the gene on the two homologous chromosomes. Thus, a monoclonal cancer, although heterogeneous genetically,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.



Abbreviation: IC, interference/codominance.

†To whom reprint requests should be addressed at: Division of Biology, 147–75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125. e-mail: avarsh@cco.caltech.edu.

always contains a set of founder mutations that is shared by all of its cells.

A drug that kills a cell only if it lacks a specific macromolecular target would distinguish tumor cells from many other cells of an organism, provided that the target is a product of a gene that had been deleted or inactivated in this cancer at the stage of its founder cell. Such a drug may be especially selective against cancers that lack a gene for a ubiquitously expressed tumor suppressor, for example, the retinoblastoma (Rb) protein (11, 12). An example of the negative-target approach is the use of a mutant adenovirus that replicates selectively in human cancer cells lacking the tumor suppressor p53 and has been shown to kill these cells in a model setting (8).

However, other tumor suppressors may not be expressed at comparable levels in most cells. A drug that kills a cell if it lacks a nonubiquitous tumor suppressor would be toxic to a subset of normal cells as well. This problem could be reduced through the use of a drug that is toxic only if a cell lacks two specific macromolecules, termed negative targets. Two judiciously chosen negative targets may, together, suffice to distinguish all of the cancer cells from all of the organism's normal cells. If they do not, a third negative target that had been deleted or rendered defective in a given cancer can be employed as well. This strategy requires a drug that possesses the ability to kill a cell if it lacks two or more of the predetermined targets, but would spare a cell containing either one of these targets.

Other changes in a founder cell may involve a missense mutation, an amplification and overexpression, an ectopic expression, or a translocation/fusion of a specific protooncogene such as, for example, *Ras* or *Myc* (9, 10, 13). A single oncoprotein may not be a unique enough target by itself, for reasons similar to those described above in the context of negative targets. However, a combination of two or more distinct oncoproteins that were either mutated or inappropriately expressed in the founder cell can be employed to formulate the unique multiprotein signature of a specific cancer that comprises both positive and negative targets.

These considerations suggest that a conditionally cytotoxic therapeutic regimen that is exquisitely specific for a given cancer, and therefore would eliminate it without significant side effects, must possess, in most cases, a multitarget, combinatorial (positive/negative) selectivity of the kind defined above. Conversely, even an informed choice of the molecular target for a single-target drug may not suffice to define unambiguously the cell type to be eliminated. Note that simply combining two single-target drugs against two different targets in a multidrug regimen would not yield a multitarget selectivity, because the two drugs together would perturb not only cells containing both targets but also cells containing either one of the targets.

Although the problem of insufficient selectivity is not as acute with noncytotoxic drugs, it is relevant to them as well. Among the multitude of examples are side effects of therapies with antipsychotic agents. The side effects are caused in part by the insufficient molecular specificity of drugs, which is exemplified by the ability of antidepressants that inhibit monoamine oxidase to perturb other proteins as well (2). This difficulty will continue to abate with the development of more specific single-target inhibitors. But an entirely distinct, major, and unsolved problem with inhibitors as drugs is the current impossibility of restricting their action to a specific subset of cells among those that contain the inhibitor's target. For example, even an exquisitely specific inhibitor of a clinically relevant enzyme is likely to have side effects, because the target enzyme is present, in most cases, not only in the cells where its inhibition is clinically beneficial but also in the cells where its inhibition is physiologically inappropriate. The present work describes a possible solution of this problem.

A previously proposed approach to designing multitarget drugs utilized degradation signals (degrons) and analogous signals that exhibit the property of codominance (14, 15). As a result, this strategy was confined to macromolecular reagents such as proteins, with the attendant problems of immunogenicity, extravasation, and intracellular delivery. The latter difficulty is especially significant, because either gene-therapy or direct-delivery methods for introducing large molecules into cells work reasonably well with cells in culture but are still inefficient with cells in an intact organism. The challenge, then, is to attain a multitarget, combinatorial selectivity in the setting of small (≤1 kDa) drugs, where the immunogenicity and delivery problems are less severe. A solution, described below, invokes a modification of the earlier idea of codominant interference (14) in conjunction with the new concept of antieffectors. This solution is applicable to either cytotoxic or noncytotoxic therapies.

## RESULTS AND DISCUSSION

**Multitarget Compounds Specific for Negative Targets: The Concept of Codominant Interference.** Previous work (14) suggested that the property of codominance, characteristic of degradation signals (degrons) and many other signals in biopolymers, can be employed to design protein-based reagents that possess multitarget, combinatorial selectivity of the kind defined above. Codominance refers to the ability of two or more signals in the same molecule to function independently and not to interfere with each other. It is shown below that a distinct version of the interference/codominance (IC) concept (14) is applicable to small (≤1 kDa) compounds. Consider a reagent containing three small moieties **a**, **b**, and **i**, which can bind, respectively, to three macromolecular targets **A**, **B**, and **I**. Because the moieties **a**, **b**, and **i** are much smaller than the macromolecules **A**, **B**, and **I**, it should be possible to arrange these moieties in the compound **abi** in such a way that the binding of **A** or **B** to **a** or **b** would preclude, through steric hindrance, the binding of moiety **i** to **I** (Fig. 1).

That the interactions of a small bipartite compound with its two macromolecular ligands can be made mutually exclusive is expected from basic physicochemical considerations. This has also been demonstrated directly, in a context unrelated to the present discussion. When lisinopril, an inhibitor of the angiotensin-converting enzyme (ACE), was connected, via an 11-atom linker, to the biotin moiety, the resulting bivalent compound could bind to and inhibit ACE in the absence but not in the presence of the biotin-binding protein streptavidin (16). Small compounds comprising two linker-connected moieties such as cyclosporin and FK506, which are specific for two macromolecular targets, have previously been employed as *in vivo* dimerization devices, making it possible to bring together two otherwise noninteracting proteins (17). However, the linker moiety of these bipartite compounds was chosen to allow simultaneous interactions with the targets, in contrast to the mutual exclusivity of interactions in the IC approach (Fig. 1).

If the moiety **i** is an inhibitor of an essential cellular enzyme **I**, the presence of the macromolecular targets **A** or **B** in a cell would reduce the inhibition of enzyme **I** by **abi**, because the complexes **abi-A** and **abi-B** would be mutually exclusive with the complex **abi-I** (Fig. 1). Note that **A** and **B** are codominant in their ability to reduce the inhibition of **I** by **abi**. Therefore, there is, formally, no limit on the number of **a**, **b**-like competition modules that can be used to construct an **abi**-like compound whose activity is sensitive to the presence of several distinct macromolecules, called negative targets. The fractional occupancy of the macromolecular targets **A**, **B**, and **I** by the **a**, **b**, and **i** moieties of **abi** would be determined in part by the targets' intracellular concentrations. There are also specific pharmacokinetic constraints on the selectivity of **abi**, an issue discussed below.

A tabulation of the relative toxicities of **abi** for cells that either lack or contain the negative targets **A** and **B** is shown in Fig. 2. It can be seen that **abi** would be relatively nontoxic to three of the four cell types and toxic exclusively to the cells that lack both **A** and **B** (Fig. 2). Thus, the IC concept allows the construction of

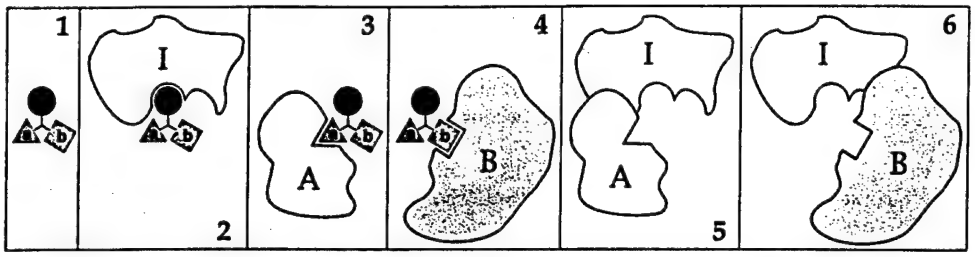



FIG. 1. The interference/codominance concept. (*1–4*) A small (<1 kDa) moiety **i** is linked to two other small moieties, **a** and **b**. The moieties **a**, **b**, and **i** are ligands of the macromolecules **A**, **B**, and **I**, respectively. The distances between **a**, **b**, and **i**, and their mutual arrangement in the tripartite compound **abi** are such that the interaction **i-I** is mutually exclusive with either the interaction **a-A** or the interaction **b-B**. Specifically, the macromolecule **I** in its complex with the small moiety **i** would sterically clash with the macromolecules **A** or **B** if either **A** or **B** is positioned to bind **a** or **b** of **abi** (*5* and *6*). In the diagram, the interactions **a-A** and **b-B** are also mutually exclusive, but this constraint is not essential. (Note that if the interactions **a-A** and **b-B** were mutually nonexclusive, the compound **abi** would promote the binding of **A** to **B**.) The codominance aspect of the IC concept allows this design to accommodate more than two of the **a**, **b**-like competition modules (not shown). In Figs. 2–4, the **i** moiety is an inhibitor of an essential enzyme **I**. In fact, the only constraint on the identities of **i** and **I** is the requirement for an **i-I** interaction to alter the functional activity of a macromolecule **I**. In other words, the choice of **I** is determined by the intended effect of the (unsequestered) compound **abi** (see the main text).

small compounds that exhibit multitarget selectivity for negative targets. One more idea is required to accomplish the same for positive targets and to link the two strategies.

**Multitarget Compounds Specific for Positive Targets: The Concept of Antieffectors.** Consider a small compound **i*** that binds to enzyme **I** in the vicinity of its active site, but does not perturb the catalytic activity of **I** toward its physiological substrates (Fig. 3*B*). Suppose further that the compound **i***, termed an antiinhibitor, was designed to interfere, sterically, with the binding of an inhibitor **i** to the enzyme's active site while at the same time allowing the binding of physiological substrates. One way to achieve this would be to endow either **i**, or **i***, or both of them with a set of chemical groups, termed a "bump," whose function is to produce steric hindrance that makes the interactions **i-I** and **i*-I** mutually exclusive (Fig. 3 *A* and *B*). A moiety that functions as a bump may also be designed to enhance specific binding of either the inhibitor **i** or the antiinhibitor **i*** to enzyme **I**, but this consideration is secondary to the bump's essential purpose.

In one application of the antiinhibitor **i***, it is linked to **c**, a small moiety that can bind to a macromolecular target **C**. The mutual arrangement of **i*** and **c** in **ci*** is such that the interactions of **ci*** with **I** and **C** are mutually exclusive. In the absence of **C**, **ci*** would compete with **abi** for the binding to enzyme **I**, thereby partially protecting **I** from inhibition by **abi** (Fig. 3*C*). This protective effect of **ci*** would be suppressed in the presence of its macromolecular target **C** (Fig. 3*D*). In the logic of codominance, discussed above in the context of negative targets, a compound bearing an antiinhibitor moiety **i*** could contain more than one **c**-like moiety. For example, a compound **cdi***, whose moieties **c** and **d** can bind, respectively, to the macromolecules **C** and **D**, would reduce the inhibition of enzyme **I** by **abi** only in the absence of both **C** and **D**. Yet another pattern of multitarget selectivity can be produced, in this context, by separating the competition moieties **c** and **d**. The resulting **ci*** and **di***, if administered together with **abi**, would reduce the inhibition of enzyme **I** by **abi** if just one of the targets, **C** or **D**, is absent. As shown below, the key merit of the antiinhibitor idea is that it allows the effect of a single inhibitor **i** to be modulated by both negative and positive macromolecular targets.

**On the Difference Between Antieffectors and Antagonists.** The distinctions between substrates, inhibitors, and antiinhibitors were described above. The concept of antieffectors is also relevant to ligand-binding biopolymers other than enzymes. For example, an agonist binds to its receptor and evokes a physiological response. An antagonist binds to a site of the receptor that overlaps with the agonist-binding site, does not

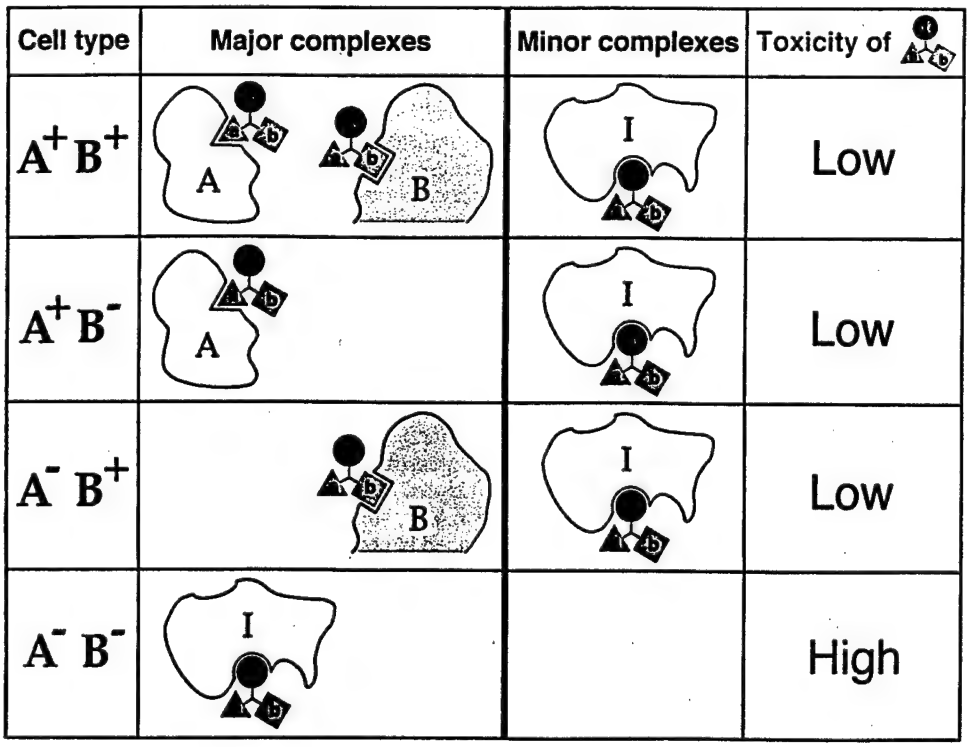



FIG. 2. Multitarget selectivity of a compound that utilizes interference/codominance. This diagram tabulates the relative toxicities of the compound **abi** for cells that either lack or contain macromolecular targets **A** and **B**. The **i** moiety of the compound **abi** (see the legend to Fig. 1) inhibits an essential enzyme **I**. The interaction **i-I** is mutually exclusive with the interaction **a-A** and the interaction **b-B**, the macromolecules **A** and **B** being negative targets of **abi**. It is assumed that concentrations of the targets **A** and **B** in cells that contain at least one of them significantly exceed the concentration of **I** (see the main text). In $A^+ B^+$, $A^+ B^-$, and $A^- B^+$ cells, the enzyme **I** would be at most partially inhibited by the **i** moiety of **abi**, because of the competing interactions of **abi** with **A** and/or **B**. By contrast, in $A^- B^-$ cells, the bulk of **abi** molecules would be available for interaction with **I**, resulting in the selective toxicity of **abi** to these cells. The selectivity pattern of **abi** requires that certain pharmacokinetic conditions are met as well (see the main text). Note that the physiological effects and the uses of **abi**-type compounds are not confined to cytotoxic regimens.

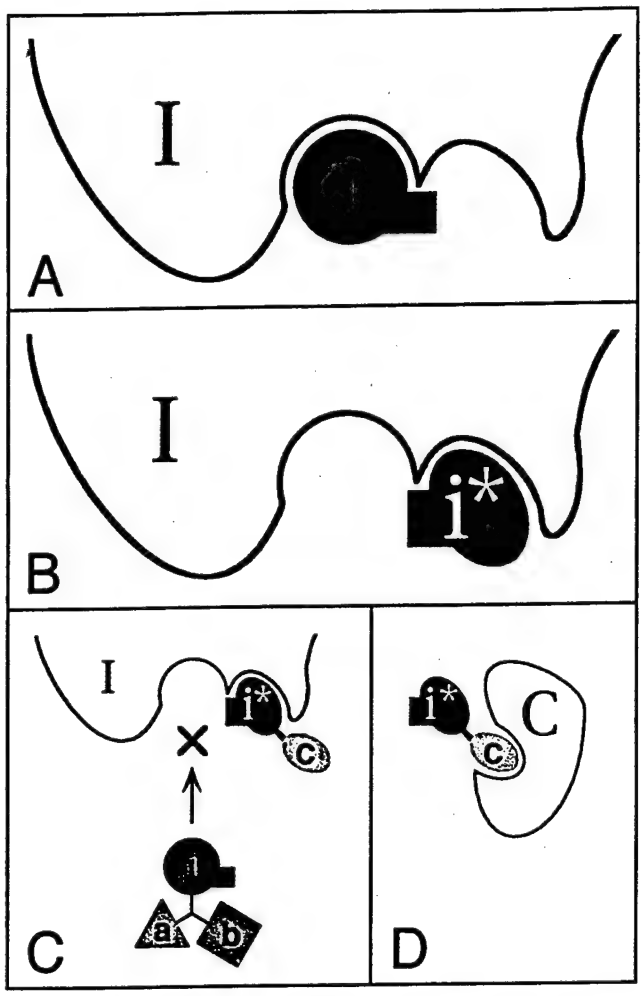



FIG. 3. The antieffector concept. The particular case illustrated here is that of an antiinhibitor **i***, defined as a compound whose binding to an enzyme **I** does not inhibit the activity of **I** but does preclude the inhibition of **I** by an inhibitor **i**. In this example, the antiinhibitor **i*** has the following properties. First, it binds to **I** in the vicinity of the **I**'s active site, but does not perturb the catalytic activity of **I** toward its physiological substrates. Second, **i***, in its bound state, sterically interferes with the interaction between **I** and its inhibitor **i**. To implement the second condition, either **i**, or **i***, or both of them bear additional moiety, a "bump," denoted by the rectangular protrusions in **i** and **i***. The function of the bump is to produce steric hindrance that makes the interactions **i-I** and **i*-I** mutually exclusive. The inhibitor **i** described here and in the main text is a competitive inhibitor, but **i** could be a noncompetitive inhibitor as well. An allosteric antiinhibitor, which functions through binding to a remote site of enzyme **I**, is yet another possibility. (*A*) A complex of the enzyme **I** with its inhibitor **i**. (*B*) A complex of **I** with its antiinhibitor **i***. Note that the bumps of the bound **i** and **i*** spatially overlap. (*C*) The antiinhibitor **i*** is linked to **c**, a small moiety that can bind to a macromolecular target **C**. The design of **ci*** is analogous to **abi** (Figs. 1 and 2), in that the interactions of **ci*** with **I** and **C** are mutually exclusive. When **ci*** is bound to **I**, the inhibitor **i**, shown here as a part of the compound **abi** (Figs. 1 and 2), is unable to bind to and inhibit the enzyme **I**. (*D*) A complex between **ci*** and its macromolecular target **C**. This complex, being mutually exclusive with the **ci*-I** complex, reduces the ability of **ci*** to protect the enzyme **I** from inhibition by **abi**.

activate the receptor, and in addition precludes the binding of agonist (1, 2). By contrast, an antieffector, which would be called, in this setting, an antiantagonist, binds to the receptor in such a way that the receptor can still bind, and respond to, the agonist, but cannot bind the antagonist. To this end, either an antagonist, or an antiantagonist, or both must possess a bump, an additional moiety described above in the context of enzymes and antiinhibitors (Fig. 3). The idea of antieffectors is thus distinct from that of antagonists or inhibitors and is new, to the best of my knowledge. I am also not aware of a naturally occurring pair of compounds that satisfy the definition of effectors/antieffectors in a physiologically relevant setting.

**Interference/Codominance and Antieffector in a Regimen That Possesses Combinatorial Selectivity.** Applying the IC and antieffector concepts together yields regimens that possess true combinatorial selectivity, i.e., sensitivity to both negative and positive targets. Consider a population of cells that either contain or lack the macromolecular targets **A**, **B**, and **C**. Our aim is to devise a treatment that would be toxic to cells that lack **A** and **B** but contain **C** ($A^-$ $B^-$ $C^+$ cells) and relatively nontoxic to the other cell types (Fig. 4). A regimen of two compounds, **abi** (Fig. 3*C*) and **ci*** (Fig. 3*D*), has the requisite selectivity, as shown in Fig. 4, which tabulates the outcomes of this treatment for different cell types. Specifically, in the absence of **C** (four cell types out of eight), the antiinhibitor-containing **ci*** would compete with the inhibitor-containing **abi** for binding to the essential enzyme **I**, thereby reducing the inhibition of **I** by **abi**, and hence reducing the toxicity of **abi**. In three other cell types, whose common property is the presence of **C** and at least one of the other two targets, **A** or **B**, the antiinhibitor-containing **ci*** would be largely sequestered by **C**, and hence inactive, but the inhibitor-containing **abi** would be sequestered as well, by either **A** or **B**. In only one type of cells, those that lack **A** and **B** but contain **C** ($A^-$ $B^-$ $C^+$ cells), is the inhibitor **abi** fully available for interaction with **I**, resulting in higher toxicity (Fig. 4). The differences in the toxicity of **abi** to different cell types would be determined by the relative stoichiometries and absolute concentrations of the cellular targets involved (**A**, **B**, **C**, and **I**), by the affinities of the moieties **a**, **b**, **c**, **i**, and **i*** for these targets, and by the pharmacokinetic properties of **abi** and **ci***.

Straightforward variations of the **abi** and **ci*** designs that utilize IC and the properties of antiinhibitors would allow selective targeting of any one of the eight cell types that differ by the presence or absence of three macromolecular targets. Moreover, there is no formal limit on the total number of negative and/or positive targets that can be simultaneously sensed by regimens that employ **abi**- and **ci***-type compounds bearing multiple interference moieties. Note that the cell type selectivity of regimens such as **abi** + **ci*** (Fig. 4) is analogous to the outcomes of the Boolean operations "and," "or," and combinations thereof.

**Stoichiometries, Affinities, and Pharmacokinetics.** The selectivity of the proposed compounds results from mutually exclusive, competing interactions between individual moieties of these compounds and their macromolecular targets (Figs. 2–4); hence, the importance of the targets' intracellular concentrations, relative to each other and the enzyme **I**, which is inhibited by an effector moiety of these drugs. The choice of **I** is not confined to essential enzymes. The target **I** could be, for instance, a DNA-binding repressor of terminal differentiation, a repressor of apoptosis, or, in the example of a noncytotoxic therapy, a neurotransmitter-inactivating enzyme. In other words, the choice of **I** is determined by the intended effect of the (unsequestered) compound **abi**.

The sequestration of **abi**- and **ci***-type compounds by their macromolecular ligands **A**, **B**, and **C** serves to prevent their binding to the enzyme **I** (Figs. 3 and 4). Therefore, in schemes of the type considered above, the molar concentration of **I** should be significantly (if possible, considerably) lower than the molar concentrations of **A**, **B**, and **C**. In addition, the concentration of **ci*** in a **ci*** + **abi** regimen should significantly exceed that of **abi**, because **ci*** is the sole obstacle to the inhibition of enzyme **I** by **abi** in the $A^-$ $B^-$ $C^-$ cells (Fig. 4). It is assumed, furthermore, that the total intracellular concentrations of **abi** and **ci***, bound and unbound, would remain significantly below the concentrations of the interference targets **A**, **B**, and **C**. The affinities of **A**, **B**, **C**, and **I** for the

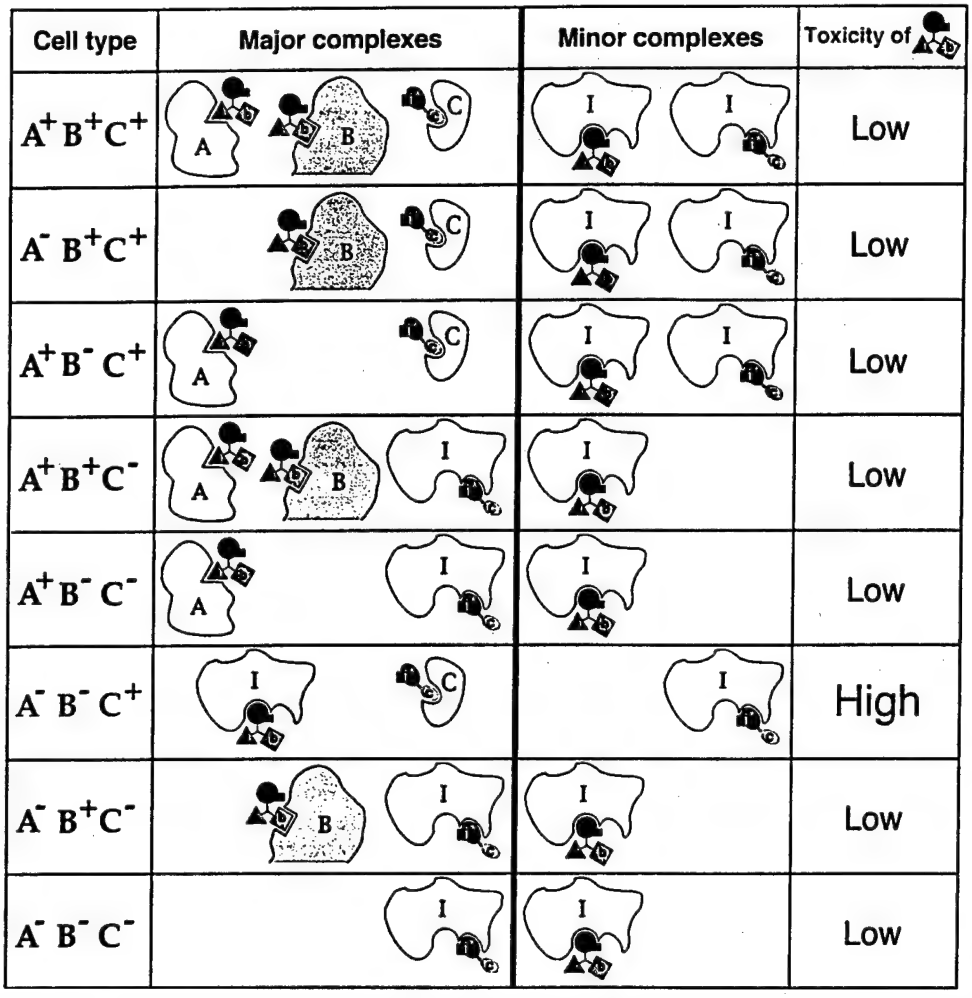



FIG. 4. Combinatorial (positive/negative) selectivity of a regimen that utilizes interference/codominance and antiinhibitor. The diagram tabulates the relative toxicities of the compound **abi** in the setting of a two-compound treatment of cells that either lack or contain the macromolecular targets **A**, **B**, and **C**. The inhibitor-containing compound **abi** is described in the main text and Figs. 1, 2, and 3*C*. The antiinhibitor-containing compound **ci*** is described in the main text and Fig. 3. **A** and **B** are negative targets, in that they reduce, through the binding to the moieties **a** and **b** of **abi**, the inhibition of an essential enzyme **I** by **abi**. **C** is a positive target, in that it reduces, through the binding to the **c** moiety of **ci***, the binding of **ci*** to enzyme **I**. This results in a larger fraction of enzyme **I** available for the inhibition by **abi**. It is assumed that the concentrations of **A**, **B**, and **C** in cells that contain them significantly exceed the concentration of **I** (see the main text). In all of the cell types except $A^-$ $B^-$ $C^+$, the enzyme **I** would be at most partially inhibited by the **i** moiety of **abi**, because of the competing interactions of **abi** with **A** and/or **B**, and also because in $C^-$ cells a fraction of enzyme **I** would be protected from the inhibition by **abi** through the interaction of **I** with the antiinhibitor moiety **i*** of **ci***. By contrast, in $A^-$ $B^-$ $C^+$ cells, the antiinhibitor-containing **ci*** would be sequestered by **C**, whereas **abi** would not be sequestered by **A** or **B**, which are absent from these cells. As a result, a larger fraction of the inhibitor-containing **abi** molecules would be available for the interaction with **I**, resulting in the selective toxicity of **abi** to $A^-$ $B^-$ $C^+$ cells. The selectivity pattern of **abi** requires that certain pharmacokinetic conditions are met as well (see the main text). Note that the physiological effects and the uses of **abi**-type compounds are not confined to cytotoxic regimens.

respective moieties of the compounds **abi** and **ci***, and also the targets' intracellular locations are among the independent parameters that can be varied in designing these compounds.

Yet another, and major, constraint on the nature and pharmacokinetics of multitarget drugs stems from the fact that the selectivity patterns described above (Figs. 2 and 4) may not be observed under equilibrium conditions, where the influx of a drug into cells equals its outflux. To illustrate one clear difficulty, let us oversimplify and suppose that an **abi**-type compound is metabolically inert, in addition to being capable of crossing the plasma membranes and other lipid bilayers. If **abi** is initially outside the cells, and if the extracellular pool of **abi** (e.g., in the blood plasma) is large enough, it can be shown that the subsequently reached equilibrium state would be characterized by equal concentrations of the free **abi** in both the $A^+$ $B^+$ and $A^-$ $B^-$ cells, thereby resulting in the equal occupancies of the enzyme **I** by **abi** in these cells, contrary to the pattern illustrated in Fig. 2.

By contrast, the selectivity patterns of Figs. 2 and 4 would be observed during the initial influx of drugs into cells. Thus, one requirement for the multitarget selectivity of **abi**- and **ci***-type regimens is the avoidance of equilibrium states such as the one described above. This and related considerations indicate that despite the logical simplicity of the proposed designs, their implementation will have to address pharmacokinetic problems that do not necessarily arise with single-target drugs.

**Selection of Targets and Construction of Multitarget Drugs.** Appropriate macromolecular targets of the **A–C** class (Figs. 3 and 4) are suggested by the protein composition of the tumor cells to be eliminated. The choice of an essential intracellular enzyme **I** (Figs. 2–4) is determined by the presence of **I** at least in tumor cells, its physiological concentration, and the feasibility of an efficacious inhibitor of **I**. Among potentially suitable enzymes for which cell-penetrating inhibitors already exist is dihydrofolate reductase. Its high-affinity inhibitors include methotrexate, which enters cells through carrier-mediated pathways, and the more lipophilic trimetrexate, which can enter cells by diffusing through lipid bilayers (18).

Each of the modules in the **abi**- and **ci***-type compounds (Figs. 1–4) would bind its macromolecular target in the absence of the other modules. Therefore, the cytotoxic **i**-type modules of IC-based compounds would be similar to the stand-alone cytotoxic drugs of today. By contrast, the interference modules of these compounds, i.e., their **a**-, **b**-, and **c**-type moieties (Figs. 1–3), are supposed to bind to their macromolecular targets but preferably not impair them functionally. This specification of a competition module simplifies its design in comparison to that of inhibitors, because many sites on the target's surface, and not just the active site, would be acceptable.

Antiinhibitors (Fig. 3) are a new class of physiologically active compounds. The opportunities and problems of their design are similar to those for the interference moieties **a–c** (Figs. 1–3), but there are two other difficulties as well. First, the antiinhibitor moiety **i*** must bind in the vicinity of, but not at, the active site of enzyme **I**. Second, the moiety **i*** must also bear chemical groups (a bump) whose function is to preclude, through steric hindrance, the binding of the inhibitor moiety **i** to the enzyme **I** (Fig. 3).

A substrate-binding cleft is not the only indentation in a folded protein molecule. The other clefts tend to be smaller, but they are present as well (19), and some of them may be located next to the enzyme's active site (Fig. 3). In addition, even relatively flat molecular surfaces can be, in principle, the sites of high-affinity interactions with small ligands (20). These optimistic comments notwithstanding, the development of antiinhibitors is certain to be a complex undertaking. As discussed above, the pharmacokinetic aspects of the proposed designs are also complex. Yet simplicity is good only if it works. Single-target anticancer drugs remain unsatisfactory, in spite of decades of immense effort. It may therefore be wise to attempt a more complex but also more effective solution.

A recent advance in drug design, termed SAR by NMR (structure–activity relationships by nuclear magnetic resonance), provides an especially promising route to constructing ligands for specific regions of a protein molecule (21). In this approach, a library of small molecules is screened for binding to an $^{15}$N-labeled protein by using NMR, which can detect weak interactions, and in addition assigns them to specific nitrogens of a protein, thereby identifying the site of binding. Finding two small compounds that bind to adjacent patches of the target protein molecule and covalently linking these compounds produces a higher-affinity ligand. This powerful strategy (21), which already yielded tightly binding ligands of specific proteins, may prove sufficient for constructing the **a**–**c** competition modules and the **i*** antiinhibitor modules of the proposed designs (Figs. 1–4).

**Noncytotoxic Multitarget Drugs.** Many useful drugs are the inhibitors of intracellular enzymes that are not essential for cell viability (2). The problem of insufficient selectivity is relevant to these drugs as well. For example, even an exquisitely specific inhibitor of a clinically relevant enzyme is likely to have significant side effects, because the target enzyme is present, in most cases, not only in the cells where its inhibition is clinically beneficial but also in the cells where its inhibition is physiologically inappropriate. The logic of **abi**-type inhibitors (Figs. 1 and 2) and **ci***-type antiinhibitors (Figs. 3 and 4) is applicable in these settings, because an informed choice of the competition moieties **a**, **b**, and **c** would sharpen up the cell selectivity of the inhibitor moiety **i** in the way described above for cytotoxic drugs (Fig. 4), resulting in the inhibition of the (nonessential) enzyme **I** in a predetermined subset of the enzyme-containing cells. Note that the same considerations apply to extracellular settings as well. The examples above are but a glimpse of the drug-engineering vistas that are opened up by the IC and antieffector concepts. At the same time, there are significant pharmacokinetic constraints on the properties of the proposed drugs, as discussed above. These constraints are likely to complicate the implementation of the IC/antieffector strategies.

**The Problem of Drug Resistance.** With small anticancer drugs that are in use today, the macromolecular target of a drug serves two distinct functions. First, the target is a cell-selectivity determinant that may bias the treatment against tumor cells. Second, the target is also a device whose inhibition by the drug brings about the desired effect, e.g., cell death. Consequently, when drug-resistant tumor cells, selected by a drug treatment, present themselves as a cancer recurrence, the necessity of employing another therapeutic agent (if such an option exists) robs the physician of whatever cell-selectivity advantage there was with the earlier drug.

The situation is qualitatively different with IC-based compounds. Suppose that a treatment that included the drug **abi** (Figs. 1 and 2) results in the appearance of **abi**-resistant tumor cells that contain, for example, an altered or overproduced enzyme **I**. If so, replacement of the **i** moiety by another small cytotoxic moiety, specific for another essential enzyme, would retain the cell selectivity of the new **ab**-containing drug. Thus, one advantage of modularity inherent in the designs of IC/antiinhibitor-based compounds (Figs. 1–4) lies in the separation of the effector aspect of a drug from its selectivity aspect. As a result, once an efficacious arrangement of the selectivity modules in **abi**- or **ci***-type compounds has been identified, it can be reutilized in drugs bearing effector moieties other than **i** and **i***.

**Concluding Remarks.** The above considerations are based on the existing understanding of single-target drugs and on the notion of steric hindrance. By introducing the new concept of antieffectors and a modification of the previously proposed idea of codominant interference (14), we can now attempt the construction of small modular compounds that possess a multitarget, combinatorial selectivity (Fig. 4). The IC/antieffector strategies are not confined to cytotoxic therapies and are relevant, in principle, to all pharmacological settings. As indicated above, one expected difficulty in implementing these strategies stems from significant pharmacokinetic constraints that do not necessarily arise with single-target drugs.

This work was motivated by the premise that the confinement of anticancer drug research and development to single-target compounds will prove insufficient for the task at hand, because even the informed choices of targets for such drugs may not define unambiguously enough the cell type to be eliminated. The remedy, described above, is to aim for drugs that possess qualitatively different selectivity—multitarget and combinatorial. If this view is correct, the future ascent of multitarget drugs may transform not only the treatment of cancer but also approaches in other settings where the killing or modification of undesirable cells or organelles is carried out in the presence of nearly identical cells or organelles that must be spared. These applications of multitarget drugs encompass more discriminating antiviral and antifungal therapies, as well as the selective killing of activated lymphocytes in autoimmune diseases and the selective elimination of damaged mitochondria in aging cells (14, 15). In yet another class of applications, a noncytotoxic multitarget drug would be used to inhibit a clinically relevant nonessential enzyme in a specific subset of the enzyme-containing cells, thereby retaining the benefits of inhibition while reducing its side effects.

I thank L. Peck, G. Turner, D. Anderson, A. Rich, S. Mayo, E. Berezutskaya, and A. Kashina for comments on the manuscript. Studies in my laboratory are supported by grants from the National Institutes of Health.

1. Gilman, A. G., Rall, T. W., Nies, A. S. & Taylor, P. (1990) *The Pharmacological Basis of Therapeutics* (Pergamon, New York).
2. Munson, P. L., Mueller, R. A. & Breese, G. R. (1996) *Principles of Pharmacology* (Chapman & Hall, New York).
3. Waldman, T., Zhang, Y., Dillehay, L., Yu, J. K., Vogelstein, B. & Williams, J. (1997) *Nat. Med.* **3**, 1034–1036.
4. Folkman, J. (1985) *Adv. Cancer Res.* **43**, 175–203.
5. Thrush, G. R., Lark, L. R., Clinchy, B. C. & Vitetta, E. S. (1996) *Annu. Rev. Immunol.* **14**, 49–71.
6. Rosenberg, S. A. (1996) *Annu. Rev. Med.* **47**, 481–491.
7. da Costa, L. T., Jen, J., He, T.-C., Chan, T. A., Kinzler, K. W. & Vogelstein, B. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 4192–4196.
8. Bischoff, J. R., Kirn, D. H., Williams, A., Heise, C., Horn, S., Muna, M., Ng, L., Nye, J. A., Sampson-Johannes, A., Fattaey, A. & McCormick, F. (1996) *Science* **274**, 373–376.
9. Bishop, J. M. (1995) *Genes Dev.* **9**, 1309–1315.
10. Kinzler, K. W. & Vogelstein, B. (1996) *Cell* **87**, 159–170.
11. Weinberg, R. A. (1995) *Cell* **81**, 323–330.
12. Knudson, A. G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10914–10921.
13. Hunter, T. (1997) *Cell* **88**, 333–346.
14. Varshavsky, A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 3663–3667.
15. Varshavsky, A. (1996) *Cold Spring Harbor Symp. Quant. Biol.* **60**, 461–478.
16. Bernstein, K. E., Welsh, S. L. & Inman, J. K. (1990) *Biochem. Biophys. Res. Commun.* **167**, 310–316.
17. Crabtree, G. R. & Schreiber, S. L. (1996) *Trends Biochem. Sci.* **21**, 418–422.
18. Takimoto, C. H. & Allegra, C. J. (1995) *Oncology* **9**, 649–659.
19. Laskowski, R. A., Luscombe, N. M., Swindells, M. B. & Thornton, J. M. (1996) *Protein Sci.* **5**, 2438–2452.
20. Mattos, C. & Ringe, D. (1996) *Nat. Biotech.* **14**, 595–599.
21. Hajduk, P. J., Meadows, R. P. & Fesik, S. W. (1997) *Science* **278**, 497–499.

MOLECULAR AND CELLULAR BIOLOGY, Jan. 1999, p. 182–193 Vol. 19, No. 1
0270-7306/99/$04.00+0


# Alternative Splicing Results in Differential Expression, Activity, and Localization of the Two Forms of Arginyl-tRNA-Protein Transferase, a Component of the N-End Rule Pathway

YONG TAE KWON, ANNA S. KASHINA, AND ALEXANDER VARSHAVSKY*

*Division of Biology, California Institute of Technology, Pasadena, California 91125*

Received 13 August 1998/Returned for modification 21 September 1998/Accepted 6 October 1998

**The N-end rule relates the in vivo half-life of a protein to the identity of its N-terminal residue. The underlying ubiquitin-dependent proteolytic system, called the N-end rule pathway, is organized hierarchically: N-terminal aspartate and glutamate (and also cysteine in metazoans) are secondary destabilizing residues, in that they function through their conjugation, by arginyl-tRNA-protein transferase (R-transferase), to arginine, a primary destabilizing residue. We isolated cDNA encoding the 516-residue mouse R-transferase, ATE1p, and found two species, termed *Ate1-1* and *Ate1-2*. The *Ate1* mRNAs are produced through a most unusual alternative splicing that retains one or the other of the two homologous 129-bp exons, which are adjacent in the mouse *Ate1* gene. Human *ATE1* also contains the alternative 129-bp exons, whereas the plant (*Arabidopsis thaliana*) and fly (*Drosophila melanogaster*) *Ate1* genes encode a single form of ATE1p. A fusion of ATE1-1p with green fluorescent protein (GFP) is present in both the nucleus and the cytosol, whereas ATE1-2p–GFP is exclusively cytosolic. Mouse ATE1-1p and ATE1-2p were examined by expressing them in *ate1Δ Saccharomyces cerevisiae* in the presence of test substrates that included Asp-βgal (β-galactosidase) and Cys-βgal. Both forms of the mouse R-transferase conferred instability on Asp-βgal (but not on Cys-βgal) through the arginylation of its N-terminal Asp, the ATE1-1p enzyme being more active than ATE1-2p. The ratio of *Ate1-1* to *Ate1-2* mRNA varies greatly among the mouse tissues; it is ~0.1 in the skeletal muscle, ~0.25 in the spleen, ~3.3 in the liver and brain, and ~10 in the testis, suggesting that the two R-transferases are functionally distinct.**

The half-lives of intracellular proteins range from a few seconds to many days. The rates of processive proteolysis are a function of the cell's physiological state and are controlled differentially for specific proteins. In particular, most of the damaged or otherwise abnormal proteins are metabolically unstable. Many other proteins, while long-lived as components of larger macromolecular structures such as ribosomes and oligomeric proteins, are metabolically unstable as free subunits. Regulatory proteins are often also short-lived in vivo, providing a way to generate their spatial gradients and to rapidly adjust their concentrations, or subunit compositions, through changes in the rate of their synthesis or degradation (20, 23, 28, 39, 44, 55).

The posttranslational conjugation of arginine (Arg) to the N termini of eukaryotic proteins was described 35 years ago (26), but the function of this modification, and of the enzyme involved, Arg-tRNA-protein transferase (R-transferase) (47), remained unknown until the discovery that the identity of N-terminal residue in a protein influences its metabolic stability (4). The resulting relation was termed the N-end rule (54). Aspartate (Asp) and glutamate (Glu), the two N-terminal residues known to be arginylated by R-transferase (47), were shown to be destabilizing residues in the N-end rule (4). It was therefore proposed (4) that the function of R-transferase is to target proteins for degradation by conjugating Arg, one of the primary destabilizing residues, to secondary destabilizing N-terminal residues (Asp and Glu in fungi; Asp, Glu, and Cys in metazoans (18) (Fig. 1). It was also proposed (4) that the analogous prokaryotic enzyme Leu, Phe-tRNA-protein transferase (L, F-transferase) (47) mediates the activity of N-terminal Arg and Lys, which, in prokaryotes, would be the secondary destabilizing residues. These conjectures were confirmed (7, 17, 45, 53).

The similar but distinct degradation signals which together give rise to the N-end rule are called the N-degrons (54, 56). In eukaryotes, an N-degron comprises two determinants: a destabilizing N-terminal residue and an internal Lys residue of a substrate (5, 22). The Lys residue is the site of formation of a substrate-linked multiubiquitin chain (11). The N-end rule pathway is thus one pathway of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, division, differentiation, and responses to stress (20, 23, 39, 55). In many of these processes, Ub acts through routes that involve the degradation of Ub-protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (9, 13, 40, 43).

(In the text that follows, names of mouse genes are in italics, with the first letter uppercase. Names of human and *Saccharomyces cerevisiae* genes are also in italics, all uppercase. If human and mouse genes are named in the same sentence, the mouse gene notation is used. Names of *S. cerevisiae* proteins are roman, with the first letter uppercase and an extra lowercase "p" at the end. Names of the corresponding mouse and human proteins are the same, except that all letters but the last "p" are uppercase. The latter usage is a modification of the existing convention (50), to facilitate simultaneous discussions of yeast, mouse, and human proteins. In some citations, the abbreviated name of a species precedes the gene's name.)

The N-end rule is organized hierarchically. In the yeast *S. cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase (Nt-amidase) (6), to the secondary destabilizing N-terminal residues Asp and Glu. The

* Corresponding author. Mailing address: Division of Biology, 147-75, Caltech, 1200 East California Blvd., Pasadena, CA 91125. Phone: (626) 395-3785. Fax: (626) 440-9821. E-mail: avarsh@cco.caltech.edu.

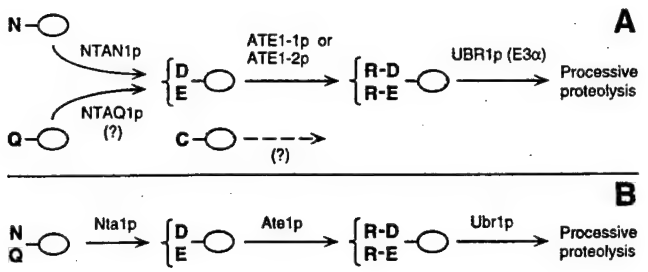



FIG. 1. Comparison of enzymatic reactions that underlie the activity of the tertiary and secondary destabilizing residues among eukaryotes. (A) Mammals (reference 54 and this work); (B) the yeast *S. cerevisiae* (5). N-terminal residues are indicated by single-letter abbreviations for amino acids; ovals denote the rest of a protein substrate. The *Ntan1*-encoded mammalian Nt-amidase converts N-terminal Asn to Asp, whereas N-terminal Gln is deamidated by a distinct Nt-amidase that remains to be identified (19, 51). In contrast, the yeast Nt-amidase can deamidate either N-terminal Asn or Gln (6). The secondary destabilizing residues Asp and Glu are arginylated by the mammalian ATE1-1p or ATE1-2p R-transferase (see Results). A Cys-specific mammalian R-transferase remains to be identified (see Results). N-terminal Arg, one of the primary destabilizing residues (54), is recognized by N-recognin (E3) (see the introduction).

destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *S. cerevisiae ATE1*-encoded Arg-tRNA-protein transferase (R-transferase), to Arg, one of the primary destabilizing residues (7) (Fig. 1B). In mammals, the deamidation step is bifurcated, in that two distinct Nt-amidases specific, respectively, for N-terminal Asn and Gln, mediate the activity of tertiary destabilizing residues (19, 51) (Fig. 1A). Mice lacking the Asn-specific Nt-amidase NTAN1p have recently been produced through targeted mutagenesis and found to be fertile, outwardly normal, but behaviorally distinct from their congenic wild-type counterparts (28a). In mammals, the set of secondary destabilizing residues contains not only Asp and Glu but also Cys, which is a stabilizing residue in yeast (18, 54) (Fig. 1).

The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded N-recognin (also called E3α), the recognition component of the N-end rule pathway (8). In *S. cerevisiae*, N-recognin is a 225-kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues, Phe, Leu, Trp, Tyr, Ile, Arg, Lys, and His (54). The *Ubr1* genes encoding mouse and human N-recognins, also called Eα (21, 41), have been cloned (29), and mouse strains lacking *Ubr1* have recently been constructed (29a).

The known functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae*, through the degradation of Cup9p, a transcriptional repressor of *PTR2* which encodes the peptide transporter (2, 10); a role in regulating the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae* (38); the degradation of Gpa1p, a Gα protein of *S. cerevisiae* (34); and the conditional degradation of alphaviral RNA polymerase in virus-infected metazoan cells (16, 54). Physiological N-end rule substrates were also identified among the proteins secreted into the host cell's cytosol by intracellular parasites such as the bacterium *Listeria monocytogenes* (46). Inhibition of the N-end rule pathway was reported to interfere with mammalian cell differentiation (24) and to delay limb regeneration in amphibians (52). Microarray-based comparisons of gene expression patterns in wild-type and congenic *ubr1*Δ strains of *S. cerevisiae* have shown that a number of yeast genes, of diverse functions, are significantly up- or down-regulated in the absence of the N-end rule pathway (39a).

The mammalian counterpart of the yeast *ATE1*-encoded R-transferase was partially purified from rabbit reticulocytes and shown to cofractionate with Arg-tRNA synthetase (12). Recent studies of the Ub-dependent proteolysis of endogenous proteins in muscle extracts suggested that the N-end rule pathway plays a major role in catabolic states that result in muscle atrophy (48, 49). A significant fraction of the N-end rule pathway's activity in muscle extracts was found to be tRNA dependent, indicating the involvement of R-transferase (48, 49). It was also reported that a crush injury to the rat sciatic nerve results in a ~10-fold increase in the rate of arginine conjugation to the N termini of unidentified proteins in the nerve's region upstream of the crush site (15, 57), suggesting an injury-induced increase in the concentration of R-transferase substrates and/or an enhanced activity of the N-end rule pathway.

In this work, we began the functional analysis of mammalian R-transferase (ATE1p) by isolating mouse cDNA encoding this enzyme. Surprisingly, we found two *Ate1* cDNA species, which were identical except for a 129-bp region that encoded similar but distinct sequences. One or the other, but not both, of the corresponding *Ate1* exons is retained in the mature *Ate1* mRNA, and the ratio of the resulting two species, *Ate1-1* and *Ate1-2*, varies greatly among mouse tissues. We also show that ATE1-1p and ATE1-2p, while differing in activity, can arginylate N-terminal Asp and Glu in model substrates. However, neither of them can arginylate N-terminal Cys, the known secondary destabilizing residue (18), suggesting the existence of a distinct Cys-specific mammalian R-transferase.

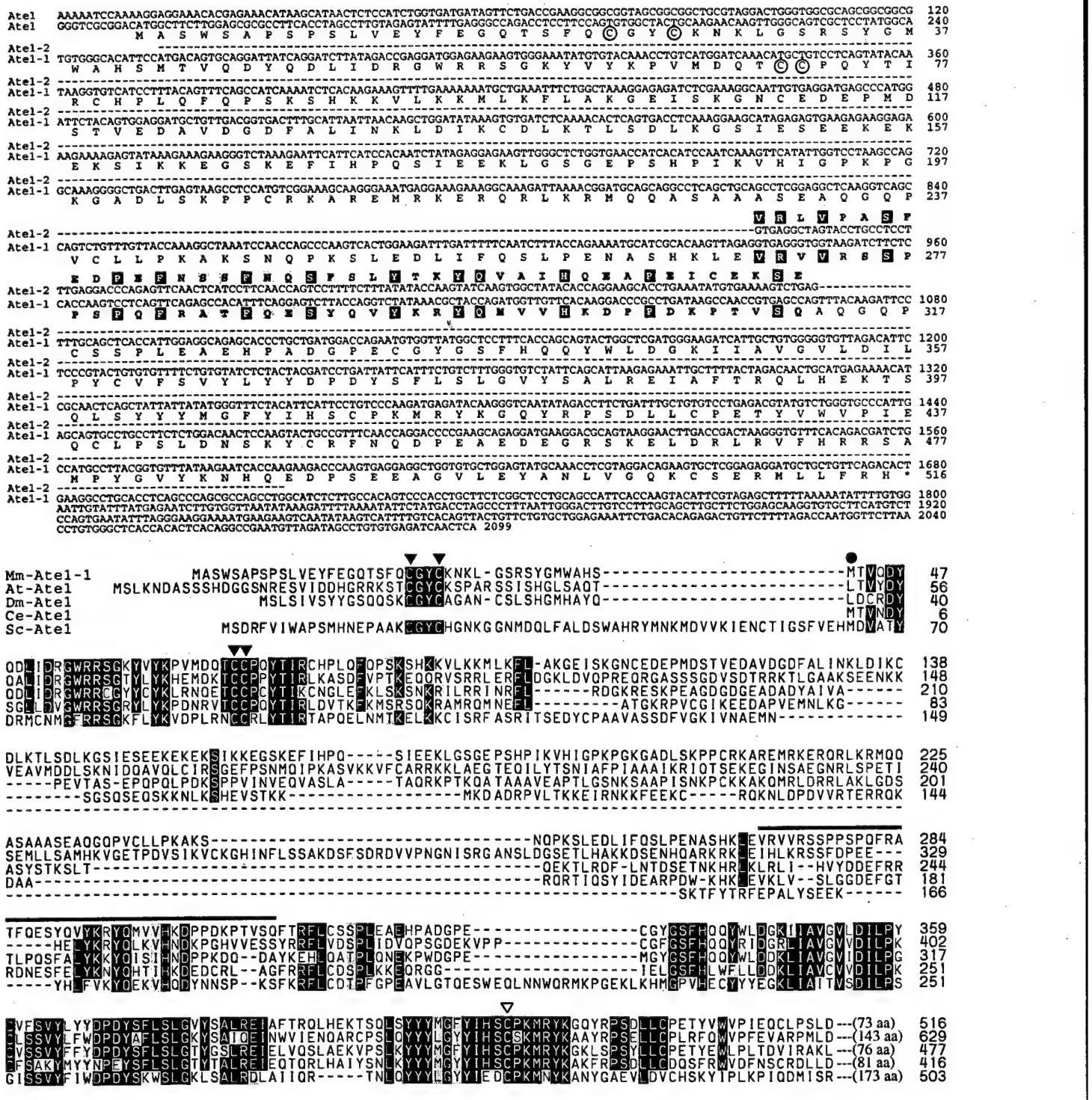
## MATERIALS AND METHODS

**Strains and plasmids.** The *S. cerevisiae* strains used were JD55 (*MAT***a** *ura3-52 his3-Δ200 leu2-3,112 trp1-Δ63 lys2-801 ubr1Δ::HIS3*) (34) and SGY3 (*MATα ura3-52 lys2-801 ade2-101 trp1-Δ63 his3-Δ200 leu2-Δ1 ate1-Δ2::LEU2*) (19a). Cells were grown in rich medium (YPD) or in synthetic media (SD) containing 0.67% yeast nitrogen base without amino acids (Difco), auxotrophic nutrients, and 2% glucose. To induce the $P_{GAL}$ promoter, glucose was replaced by 2% galactose (SG medium). Transformation of *S. cerevisiae* was performed by the lithium acetate method (3).

The *ubr1*Δ *ate1*Δ double mutant AVY34 was constructed by replacing 93% of the *ATE1* open reading frame (ORF) (the first 470 codons) in strain JD55 (*ubr1*Δ) by the *LEU2* gene, through homologous recombination (42) with the introduced *LEU2* gene flanked on either side by 40 bp of *ATE1*-specific sequences. Mutants were selected on SD (lacking Leu and His) plates, and $Leu^+$ isolates were checked by PCR for the absence of *ATE1* and by colony assays (7, 8) for the absence of Ate1p and Ubr1p activity. High-copy-number pUB23-X plasmids expressing Ub-X-βgal proteins (see below) from $P_{GAL}$ in *S. cerevisiae* have been described elsewhere (4). Mouse *Ate1-1* and *Ate1-2* cDNAs (see below) were subcloned into the low-copy-number vector p414GAL1 (36), using the engineered *Bam*HI (5′) and *Xho*I (3′) restriction sites, yielding plasmids pAT1 and pAT2. For localization assays with green fluorescent protein (GFP), cDNAs encoding mouse ATE1-1p or ATE1-2p were subcloned into the pEGFP-N1 N-terminal protein fusion vector (Clontech, Palo Alto, Calif.), using the engineered *Xho*I (5′) and *Age*I (3′) restriction sites, yielding plasmids pAT1-GFP and pAT2-GFP.

**Isolation of the mouse *Ate1-1* and *Ate1-2* cDNAs.** The 392-bp fragment of the mouse EST (expressed sequence tag) clone (accession no. AA415294), which was identified in GenBank through species walking (see Results), was used as a probe to screen a λgt10-based mouse cDNA library from MEL-C19 cells (Clontech), using standard procedures (3). Eight positive clones, whose inserts ranged from 0.5 to 1.6 kb, were analyzed by PCR and partial sequencing. The cDNA inserts of clones 3 and 8 were then subcloned into pBluescript II $SK^+$ (29) and sequenced on both strands. The resulting ORFs were identical except for a 129-bp internal region (see Results) (Fig. 2A). The deduced amino acid sequences of the mouse cDNA clones 3 and 8 were weakly but significantly similar to the deduced sequences of *Caenorhabditis elegans* and *S. cerevisiae* ATE1p (see Results) (Fig. 2B) and corresponded to nucleotides (nt) 699 to 1870 and 587 to 2099, respectively, in the subsequently produced full-length mouse *Ate1-1* and *Ate1-2* cDNAs (accession no. AF079096 and AF079097).

A human EST clone (accession no. AA503372) whose deduced amino acid sequence was highly similar to that of the partial mouse *Ate1* cDNA clone 8 was found in GenBank, using the partial mouse ATE1p sequence as a query. This EST clone was purchased from Genome Systems (St. Louis, Mo.) and sequenced


A
B
C
identity / similarity
M8
M3

to obtain more of the 5′-proximal human *ATE1* sequence. Reverse transcription-PCR (RT-PCR) (3) was then carried out with poly(A)$^+$ RNA from mouse embryonic fibroblasts, using the mouse *Ate1*-specific reverse primer 5′-CCTTT GGTAACAAACAGACTGGCTG-3′ and the forward primer 5′-TCTCATAG ACCGAGGATGGCGAAG-3′, whose sequence was derived from the above human EST clone. The resulting PCR products, which appeared as a smear upon agarose gel electrophoresis, were ligated into the TA cloning vector (Invitrogen, San Diego, Calif.), and the ligation mixture was used as a template for PCR using a nested mouse *Ate1* primer 5′-CTGCAGCTGAGGCCTGCTGCATCCG-3′ and a vector-specific primer 5′-GTTTTCCCAGTCACGAC-3′. This strategy yielded a single major DNA species (data not shown). We then applied 5′-RACE (rapid amplification of 5′ cDNA ends) (3), using the above RT-PCR-derived sequence, to produce the full-length *Ate1* cDNA as previously described (19).

**Analysis of the mouse *Ate1* gene.** The mouse genomic DNA from L cells was used as a template for PCR, using the Expand high-fidelity PCR system (Boehringer, Indianapolis, Ind.) and exon-specific primers as previously described (29), to produce DNA fragments that together spanned ~4 kb of the mouse *Ate1* gene and contained the two alternative 129-bp *Ate1* exons. The regions encompassing exon/intron junctions were sequenced by using exon- and intron-specific primers. Thereafter, a strategy described earlier for the *Ubr1* gene (29) was used to screen, using a fragment of the mouse *Ate1* cDNA (nt 255-1139), a BAC (bacterial artificial chromosome)-based library of mouse genomic DNA fragments from strain 129SvJ (Genome Systems), yielding one BAC clone containing the mouse *Ate1* gene.

**Isolation of the human, plant, and fly *ATE1* cDNAs.** Using the cloned mouse ATE1p sequences (see above) as queries, we identified in GenBank several significantly similar EST sequences from other organisms (data not shown). To determine whether these species also contained the two forms of *ATE1* mRNA, we isolated the corresponding *ATE1* cDNAs. RT-PCR (3) with poly(A)$^+$ RNA from human 293 cells and the primers 5′-CAATGGCATGTGGGCACATTCC ATG-3′ (specific for the human EST clone AA503372 [see above]) and 5′-CC ACAGGTACTGAATATGTATCCTG-3′ (specific for the human EST clone AA195361) was carried out, yielding a 1.6-kb human *ATE1* cDNA fragment lacking the first 41 codons of the *ATE1* ORF. This fragment (a mixture of the two alternative cDNAs) was subcloned into the TA vector (Invitrogen) and sequenced on both strands. Full-length *Ate1* cDNAs from *Arabidopsis thaliana* and *Drosophila melanogaster* were isolated by RT-PCR as well, using total RNA from *A. thaliana* leaves, poly(A)$^+$ RNA from *D. melanogaster* embryos, and primers specific for the 5′ and 3′ ends of the corresponding ORFs. By using the strategy described above for the human *ATE1* cDNAs, the sequences of these primers were derived from the EST clones that were initially identified in GenBank through their similarity to the mouse ATE1p sequence, then purchased from Genome Systems, and sequenced prior to RT-PCR with the corresponding RNA preparations. The final human, plant, and fly *ATE1* cDNAs were sequenced on both strands.

**Assays of β-gal.** Colony assays for the *Escherichia coli* β-galactosidase (βgal) in *S. cerevisiae* were carried out by overlaying yeast colonies on SG plates with 0.5% agarose containing 0.1% sodium dodecyl sulfate (SDS), 4% dimethylformamide, and a 0.1-mg/ml solution of the chromogenic βgal substrate X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; Calbiochem, La Jolla, Calif.), followed by incubation for 1 to 2 h at 37°C. Quantitative assays for βgal in *S. cerevisiae* were carried out with whole-cell extracts, using another chromogenic βgal substrate, *o*-nitrophenyl-β-D-galactopyranoside (ONPG). Cells in a 5-ml culture ($A_{600}$ of ~1) were pelleted by centrifugation and resuspended in 5 ml of buffer Z (60 mM $Na_2HPO_4$, 40 mM $NaH_2PO_4$, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol [pH 7.0]). After the $A_{600}$ of the suspension was determined 50- or 100-μl samples were diluted to 1 ml with buffer Z; 0.1% SDS (20 μl) and $CHCl_3$ (50 μl) were then added; the suspension was vortexed for 10 to 15 s and incubated for 15 min at 30°C, followed by the addition of 200 μl of ONPG (4 mg/ml in buffer Z) and further incubation at 30°C, until a medium yellow color had developed, at which point the reaction was stopped by the addition of 1 M $Na_2CO_3$ (0.4 ml). The mixture was centrifuged for 5 min at 1,100 × *g*, and the $A_{420}$ and $A_{500}$ of the samples were measured. The ONPG units ($U_{ONPG}$) of βgal activity were calculated as follows: $U_{ONPG} = 1{,}000 \times [(A_{420}) - (1.75 \times A_{500})]/t \times v \times A_{600}$, where $t$ and $v$ were, respectively, the time of incubation (minutes) and the sample volume (milliliters) (3).

**Purification and N-terminal sequencing of X-βgal proteins.** Extracts were prepared (using the liquid nitrogen procedure [3]) from *S. cerevisiae* AVY34 (*ubr1*Δ *ate1*Δ) cotransformed with a pUB23-X plasmid (4) (expressing Ub-X-βgal) and either pAT1 (expressing mouse ATE1-1p) or pAT2 (expressing mouse ATE1-2p). Cultures were grown in SG to an $A_{600}$ of ~1. Specific X-βgal proteins (X = Asp or Cys) were purified by affinity chromatography on ProtoSorb lacZ (Promega, Madison, Wis.), a monoclonal anti-βgal antibody coupled to agarose beads (Promega). X-βgal proteins were further purified by electrophoresis on SDS–7% polyacrylamide gels and were electroblotted onto Immobilon-P$^{SQ}$ membranes (Millipore, Bedford, Mass.). N-terminal sequencing of 10 to 15 pmol of electroblotted X-βgal was carried out for at least five cycles, using an Applied Biosystems 476A protein sequencer (Caltech Microchemistry Facility).

**Mouse cell cultures, transfection, and GFP localization.** NIH 3T3 cells (ATCC 1658-CRL) were grown as monolayers in Dulbecco's modified Eagle medium (GIBCO, Frederick, Md.) supplemented with 10% fetal bovine serum. Cells for GFP localization analyses were grown to ~15% confluence on glass coverslips for 24 h prior to transfection with either pAT1-GFP or pAT2-GFP, using Lipofectamine (GIBCO) and the manufacturer-supplied protocol. Cells were incubated for 5 h at 37°C in serum-free medium containing DNA and Lipofectamine. Thereafter an equal volume of medium containing 20% serum was added, and the cells were grown for another 12 to 20 h at 37°C. Cells were fixed with 2% formaldehyde in phosphate-buffered saline, and GFP fluorescence was visualized in a Zeiss Axiophot microscope.

**Northern hybridization.** Mouse multiple-tissue Northern blots containing 2 μg of poly(A)$^+$ RNA per lane (Clontech) were probed with the $^{32}$P-labeled 1.1-kb mouse *Ate1* cDNA (nt 638 to 1734), using the manufacturer-supplied protocol.

**Determination of the relative levels of *Ate1-1* and *Ate1-2* mRNAs.** Samples of total RNA isolated as described previously (3) from mouse spleen, skeletal muscle, liver, brain, testis, and embryonic fibroblasts were subjected to RT-PCR (28 cycles). The primers 5′-CAGTGGAGGATGCTGTTGACGGTGAC-3′ and 5′-GTGCTCTGCCTCCAATGGTGAGCTG-3′ were specific for the identical regions of *Ate1-1* and *Ate1-2* cDNAs that flanked the two 129-bp exons (see Results) which distinguished these cDNAs. The resulting 624-bp product (a mixture of the *Ate1-1* and *Ate1-2* cDNA fragments) was treated with *Scr*FI, which cuts at different sites within the two 129-bp exons, followed by a 2% agarose gel electrophoresis. This procedure made it possible to distinguish the *Ate1-1* and *Ate1-2* fragments. The ratios of the two forms of *Ate1* cDNA were determined by serial dilutions of the samples prior to gel electrophoresis.

**Nucleotide sequence accession numbers.** The nucleotide sequences reported in this paper were submitted to the GenBank/EMBL data bank and assigned accession no. AF079096 (mouse *Ate1-1* cDNA), AF079097 (mouse *Ate1-2* cDNA), AF079098 (human *Ate1-1* cDNA), AF079099 (human *ATE1-2* cDNA), AF079100 (*A. thaliana Ate1* cDNA), and AF079101 (*D. melanogaster Ate1* cDNA).

## RESULTS

**Identification of mouse *Ate1* cDNAs by species walking.** On the assumption that the sequences of R-transferases in different species might be sufficiently conserved to be detected by using the sequence of the only cloned R-transferase, *S. cerevisiae* Ate1p (7), we have been searching GenBank and related databases. No mammalian sequences in GenBank, including the EST sequences, had significant similarities to *S. cerevisiae* Ate1p. However, we did identify a nematode (*C. elegans*) ORF (accession no. Z21146) that exhibited similarity to yeast Ate1p (Fig. 2B) and then used the *C. elegans* sequence to identify a

FIG. 2. Two forms of the mouse *Ate1* cDNA and the ATE protein family. (A) The mouse *Ate1-1* and *Ate1-2* cDNAs and their products. The nucleotide sequences of *Ate1-2* identical to those of *Ate1-1* (everywhere except for the 129-bp region) are indicated by dashes. In the region of the alternative 129-bp exons of *Ate1-1* and *Ate1-2*, white-on-black and gray shadings highlight, respectively, identical and similar residues. The circled Cys residues are homologous to those that are important for the enzymatic activity of *S. cerevisiae* Ate1p (32). (B) The ATE protein family and the origins of the alternative 129-bp exons. Alignment of the sequences of mouse ATE1-1p (Mm-Ate1-1), *A. thaliana* Ate1p (At-Ate1), *D. melanogaster* Ate1p (Dm-Ate1), *C. elegans* Ate1p (Ce-Ate1), and *S. cerevisiae* Ate1p (Sc-Ate1) (accession no. J05404). Similar residues (gray) were grouped as follows: M, L, I, and V; D, E, N, and Q; R, K, and H; Y, F, and W; S, A, and T. The region encoded by the alternative 129-bp exons of mouse *Ate1* is highlighted by a thick line. Of the Cys residues that are conserved among all ATE proteins, the ones required and not required for the enzymatic activity of *S. cerevisiae* Ate1p (32) are indicated, respectively, by ▼ and ▽. The N-terminally truncated mouse ATE1-1p and ATE1-2p proteins which began at Met-42 (●) lacked the R-transferase activity (data not shown). The highly variable C-terminal regions of ATE proteins were omitted from the alignment. The sequences were aligned using PileUp program (Wisconsin Package; Genetics Computer Group, Madison, Wis.). Gaps (−) were introduced to optimize the alignment. The residue numbers are on the right of the sequences. The sequence of *C. elegans* Ate1p appears to lack the N-terminal region of other ATE proteins because of an error in defining the *Ate1* ORF in the genomic DNA sequence (accession no. Z21146). (C) Alignment of the 43-residue regions that are encoded by the alternative 129-bp exons in mammalian *Ate1*. Sequences shown: mouse (Mm-Ate1-1 and Mm-Ate1-2), human (Hs-Ate1-1 and Hs-Ate1-2), *D. melanogaster* (Dm-Ate1), *C. elegans* (Ce-Ate1), *A. thaliana* (At-Ate1), *S. pombe* (Sp-Ate1; accession no. Z99568), and *S. cerevisiae* (Sc-Ate1) (accession no. J05404). The degrees of identity and similarity of ATE proteins to the deduced amino acid sequences of the M8 (mouse ATE1-1p) or M3 (mouse ATE1-2p) exon of the mouse *Ate1* gene are indicated on the right. The residues conserved among all of the compared sequences are indicated above the alignment.

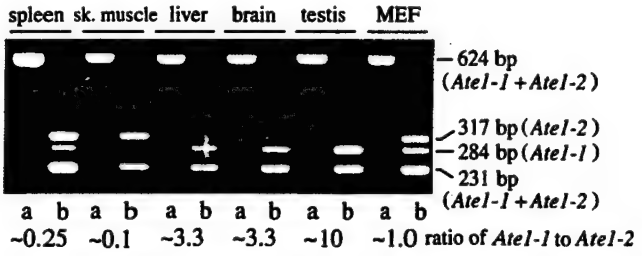



FIG. 3. Expression of *Ate1-1* and *Ate1-2* mRNAs in different mouse tissues and mouse embryonic fibroblasts (MEF). Two forms of *Ate1* cDNA (Fig. 2A) were amplified in a single reaction by RT-PCR, using the same primers, to yield 624-bp fragments that included the region of the alternative 129-bp exons (see Materials and Methods). The 624-bp fragments were digested with *Scr*FI, which produced a 231-bp fragment (a mixture of *Ate1-1* and *Ate1-2*), a 284-bp, *Ate1-1*-specific fragment, and a 317-bp, *Ate1-2*-specific fragment. The untreated (lanes a) and *Scr*FI-treated (lanes b) samples from different mouse tissues were analyzed by electrophoresis in a 2% agarose gel. The ratio of the two forms of *Ate1* mRNA, defined as the ratio of the 284-bp (*Ate1-1*) fragment to the 317-bp (*Ate1-2*) fragment, was determined by analyzing serially diluted samples and comparing the resulting band intensities (data not shown). sk., skeletal.

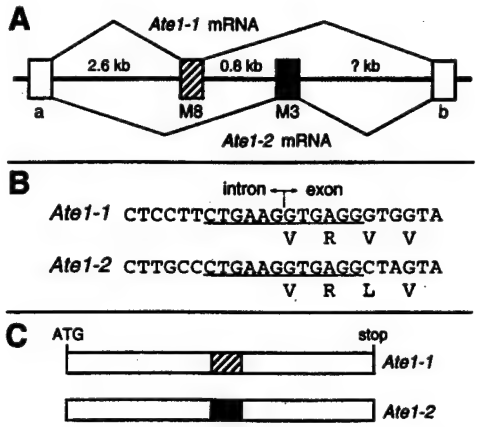



FIG. 4. The two forms of mouse *Ate1* mRNA are produced by alternative splicing. (A) The two alternative 129-bp exons are adjacent in the mouse *Ate1* gene. The thick line denotes genomic DNA; the striped and black rectangles denote the alternative 129-bp exons, M8 and M3 (see Materials and Methods); gray rectangles denote the flanking *Ate1* exons, of unknown sizes; thin lines denote the alternative splicing patterns that yield the two forms of *Ate1* mRNA. (B) The underlined 12 bp (6 bp in the intron and 6 bp in the exon) around the splice acceptor sites are identical between the two alternative 129-bp exons. (C) Scale diagrams of the two forms of mouse *Ate1* cDNAs. The alternative 129-bp exons M8 and M3 are indicated by the striped and black boxes, respectively.

significantly similar EST sequence of *D. melanogaster* cDNA (accession no. AA391570). Finally, using the deduced amino acid sequence of the *Drosophila* EST clone as a probe, we identified a 467-bp mouse EST sequence (accession no. AA415294) that exhibited weak but significant similarity to the *Drosophila* sequence but no detectable similarity to *S. cerevisiae* Ate1p. On a chance that this 467-bp EST had been derived from the mouse *Ate1* cDNA, we used it to screen a mouse cDNA library and indeed isolated the putative mouse *Ate1* cDNAs (Fig. 2A).

**Alternative splicing results in two species of mouse *Ate1* cDNA containing distinct but homologous exons.** During the initial mouse cDNA library screening, we found that the cDNA clone 3 (nt 699 to 1870 of *Ate1-2* cDNA) was identical to the cDNA clone 8 (nt 587 to 2099 of *Ate1-1* cDNA), except for a 129-bp region whose deduced amino acid sequences were similar (31% identity; 61% similarity) (Fig. 2). The two full-length *Ate1* cDNAs (termed *Ate1-1* and *Ate1-2*), which were obtained by RT-PCR followed by 5′-RACE (see Materials and Methods), encoded proteins of identical length, 516 residues (59.2 kDa and pI of 8.14 versus 59.1 kDa and pI of 7.22) (Fig. 2A), that contained regions of similarity to the 57.8-kDa Ate1p of *S. cerevisiae* (Fig. 2B). RT-PCR (followed by subcloning) with RNAs from different mouse tissues also produced the two forms of *Ate1* cDNAs, indicating that the two species were in fact present in the initial RNA preparation (Fig. 3 and data not shown).

To determine whether both of the two 129-bp regions of the *Ate1-1* and *Ate1-2* cDNAs were a part of the *Ate1* gene, and whether *Ate1-1* and *Ate1-2* were produced through alternative splicing, we analyzed the mouse *Ate1* gene in the vicinity of its two 129-bp exons, using at first PCR and subsequently a BAC clone containing *Ate1* (see Materials and Methods). The two 129-bp exons were located next to each other in the *Ate1* gene (Fig. 4A). We also found that the 12-bp sequences around the splice acceptor sites of these exons (6 bp in the intron and 6 bp in the exon) were identical between the two exons (Fig. 4B), consistent with the alternative presence of these exons in the mature *Ate1* mRNA. The exon-containing RT-PCR products from different mouse tissues appeared as a single major band retaining one of the two 129-bp exons (Fig. 3 and data not shown). Subcloning and analyses of these RT-PCR products yielded no other differentially spliced *Ate1* cDNAs (for example, cDNAs retaining both or neither of the two 129-bp exons), suggesting that the splicing of *Ate1* pre-mRNA is tightly regulated to retain one and only one of the two alternative exons. Thus, the two forms of *Ate1* mRNAs are produced by a nearly unprecedented (see Discussion) splicing pathway which ultimately yields two proteins of identical size that bear two alternative, homologous but distinct 43-residue internal sequences (Fig. 4C).

**The absence of alternative 129-bp exons from the plant and fly *Ate1* genes.** To explore the evolution of *Ate1*, and especially the phylogeny of its alternative 129-bp exons, we cloned the human, plant (*A. thaliana*), and fly (*D. melanogaster*) *ATE1* cDNAs (see Materials and Methods). Two forms of the human *ATE1* cDNA, termed *Hs-ATE1-1* and *Hs-ATE1-2*, were isolated from human 293 cells (the forms' molar ratio was about 1). However, only one form of the *Ate1* cDNA was isolated from either the leaves of *A. thaliana* (termed *At-Ate1*) or *D. melanogaster* embryos (termed *Dm-Ate1*), suggesting that the alternative 129-bp exons may not be present in the *Ate1* genes of plants and arthropods. The *A. thaliana* and *D. melanogaster* Ate1p proteins were, respectively, 629 and 477 residues long (71 and 55 kDa, with pIs of 6.0 and 8.4). Mouse ATE1-1p was 82, 38, and 42% identical (as well as 91, 57, and 61% similar) to human ATE1-1p, *A. thaliana* Ate1p, and *D. melanogaster* Ate1p, respectively (Fig. 2 and data not shown). *A. thaliana* Ate1p bore a 16-residue region containing exclusively Asp or Glu (data not shown).

We used RT-PCR and RNA preparations from *A. thaliana* and *D. melanogaster* to amplify the relevant regions of the corresponding *Ate1* cDNAs. The resulting fragments were digested with restriction enzymes that recognize, in each species, exclusively the region that corresponds to the 129-bp exons of the mouse *Ate1* cDNAs, and the products were analyzed by gel electrophoresis. The initial cDNA fragments of *A. thaliana* and *D. melanogaster Ate1* completely disappeared after this treatment, in contrast to the homologous mouse cDNA fragment (which contained two distinct sequences of identical length), suggesting that the two alternative exons were absent from the *Ate1* genes of plants and arthropods (Fig. 5A).

While this analysis was under way, complete sequences of

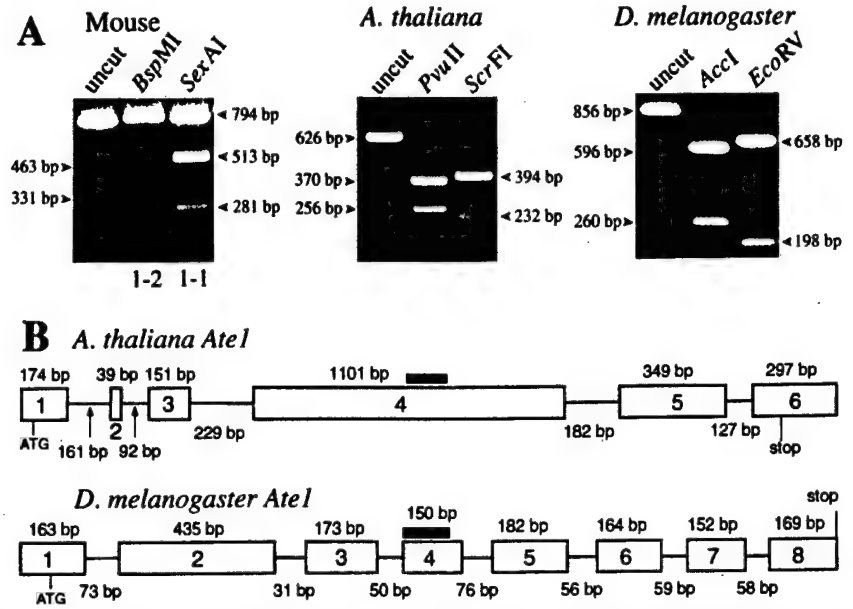



FIG. 5. Alternative splicing of *Ate1* pre-mRNA in mammals (the mouse) but in neither plants (*A. thaliana*) nor arthropods (*D. melanogaster*). (A) The relevant *Ate1* cDNA fragments from mouse (794 bp), *A. thaliana* (626 bp), and *D. melanogaster* (856 bp) were produced by RT-PCR (see Materials and Methods). The products were treated with the indicated restriction endonucleases that cut exclusively within the two alternative 129-bp exons of mouse *Ate1* cDNAs (*Bsp*MI for *Ate1-2*; *Sex*AI for *Ate1-1*) or within the corresponding regions of *A. thaliana* (*Pvu*II and *Scr*FI) and *D. melanogaster* (*Acc*I and *Eco*RV) *Ate1* cDNAs. (B) The *A. thaliana* and *D. melanogaster Ate1* genes. The exon-intron organization of these genes was deduced through comparisons of their cDNA sequences, determined in this work (see Materials and Methods), with the concurrently determined sequences of the corresponding genomic DNA regions (see text). The horizontal lines and rectangles denote, respectively, introns and exons, whose lengths are indicated below and above the line denoting introns. Thick horizontal lines indicate the regions of *A. thaliana* and *D. melanogaster* cDNAs that correspond to the alternative 129-bp exons of the mouse and human *Ate1* cDNAs (Fig. 2 and 4). The lengths of the *A. thaliana* and *D. melanogaster Ate1* genes are, respectively, ~3 and ~2.5 kb.

the *A. thaliana* and *D. melanogaster Ate1* loci, determined through the corresponding sequencing projects, were deposited in GenBank (accession no. AA005237 and accession no. AC004321, respectively). By comparing the cloned *Ate1* cDNAs (see Materials and Methods) and the corresponding genomic sequences of *A. thaliana* and *D. melanogaster*, we could deduce the organization of these *Ate1* genes. The results (Fig. 5B) directly confirmed the absence of the alternative homologous exons from *Ate1* of *A. thaliana* and *D. melanogaster*, in contrast to mammalian *Ate1*. The corresponding region of plant Ate1p is more similar to the exon-encoded sequence of mouse ATE1-1p, whereas in *Drosophila* this region is more similar to the alternative sequence of ATE1-2p (Fig. 2C). The corresponding regions of *S. cerevisiae* and *Schizosaccharomyces pombe* Ate1p are not preferentially similar to either of the two alternative exon-encoding sequences of mouse ATE1p (Fig. 2C).

**Mouse ATE1-1p and ATE1-2p can implement the Asp/Glu-specific subset of the N-end rule pathway but differ in activity.** To determine whether the two putative mouse R-transferases are in fact R-transferases and to compare their activities in an in vivo setting, we examined whether ATE1-1p and ATE1-2p could confer metabolic instability on Asp-βgal and Glu-βgal in *ate1Δ S. cerevisiae*. Asp and Glu are secondary destabilizing residues in the N-end rule (Fig. 1 and introduction). The test substrates Asp-βgal and Glu-βgal (produced through cotranslational deubiquitylation of Ub-Asp-βgal and Ub-Glu-βgal (4)) are short-lived in wild-type yeast (half-lives of ~3 and ~30 min, respectively) but long-lived (half-life of >20 h) in *ate1Δ S. cerevisiae* that lacks the *ATE1*-encoded yeast R-transferase (5, 7). Previous work (19, 33) has shown that the steady-state level of an X-βgal protein is a sensitive measure of its metabolic stability.

*S. cerevisiae ate1Δ* cells were cotransformed with a pair of plasmids that expressed one of the two putative mouse R-transferases, ATE1-1p or ATE1-2p, and one of several test substrates (as the corresponding Ub fusions): Asp-βgal, Glu-βgal, Arg-βgal, Cys-βgal, or Met-βgal. Met and Cys are stabilizing residues in the yeast N-end rule; Arg is a primary destabilizing residue; Asp and Glu are secondary destabilizing residues (5, 56). Control tests included either the vector alone or a plasmid expressing *S. cerevisiae* Ate1p. The steady-state levels of X-βgal proteins were determined by measuring the enzymatic activity of βgal in yeast extracts. Using this assay, we found that both forms of mouse ATE1p were able to confer metabolic instability on either Asp-βgal or Glu-βgal in *ate1Δ S. cerevisiae* (Fig. 6A). ATE1-1p and ATE1-2p destabilized Glu-βgal much less than Asp-βgal (Fig. 6A), consistent with Glu being a less destabilizing residue in the N-end rule than Asp, presumably because of less efficient arginylation of the N-terminal Glu by R-transferases (54). However, while the apparent destabilizing activity of the mouse ATE1-1p R-transferase was only slightly lower than that of *S. cerevisiae* Ate1p (expressed from the identical vector and promoter), the activity of mouse ATE1-2p was significantly lower than that of ATE1-1p (Fig. 6A).

We also asked whether the two forms of mouse ATE1p could influence each other's activity if they were coexpressed in the same cell (such an influence might be expected, for instance, if the active form of R-transferase were a dimer or if the two forms of R-transferase competed for binding to the same component of a pathway). *S. cerevisiae ate1Δ* cells were cotransformed with two plasmids bearing different selectable markers and expressing different combinations of ATE1-1p and ATE1-2p (1+1, 2+2, or 1+2), and also with a plasmid expressing one of the X-βgal test proteins (X = Met, Arg, Cys, Asp, or Glu). Control cells were cotransformed with the two

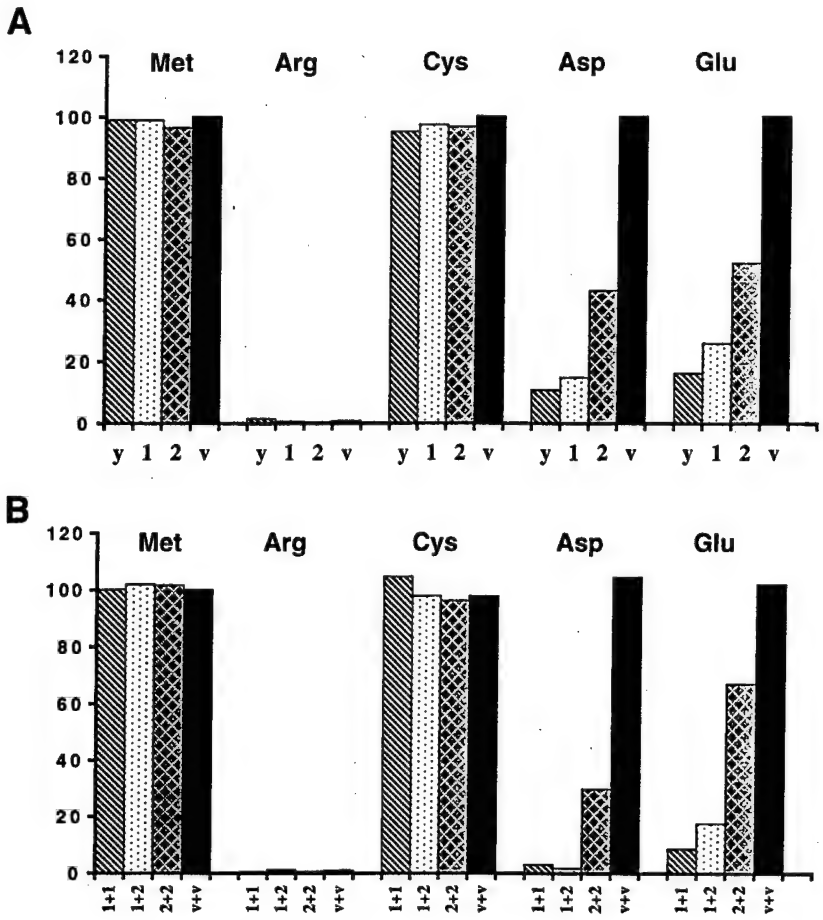



FIG. 6. The two forms of mouse ATE1p can implement the Asp/Glu-specific subset of the N-end rule pathway. (A) Relative enzymatic activities of βgal in *ate1Δ S. cerevisiae* transformed with plasmids expressing X-βgal (as Ub-X-βgal) test proteins (X = Met, Arg, Cys, Asp, or Glu) together with a plasmid expressing either yeast *ATE1* (denoted as y), mouse ATE1-1p (denoted as 1), mouse ATE1-2p (denoted as 2), or the vector alone (denoted as v). The N-terminal residues of X-βgals in each set of experiments are indicated at the top. The activity of Met-βgal in cells transformed with vector alone is taken as 100%. (B) The two forms of mouse ATE1p exhibit no cooperativity in mediating the degradation of X-βgals. Shown are relative enzymatic activities of X-βgals in *ate1Δ S. cerevisiae* strain cotransformed with plasmids expressing X-βgals (Ub-X-βgals) (X = Met, Arg, Cys, Asp, or Glu) and the combinations of plasmids expressing the following proteins: mouse ATE1-1p and ATE1-1p (1+1), ATE1-1p and ATE1-2p (1+2), ATE1-2p and ATE1-2p (2+2), or two vector controls (v+v). One of the two vectors bore the *TRP1* marker and the other bore the *HIS3* marker (see Materials and Methods). Results are averages of four independent measurements, which differed by less than 10%.

vectors alone. The results (Fig. 6B) indicated that the total activities of the 1+1 and 1+2 combinations (measured as the extent of destabilization of Asp-βgal or Glu-βgal) were similar to each other and much higher than the total activities of 2+2 (Fig. 6B), consistent with the conjecture that the two forms of mouse R-transferase do not interact and that ATE1-1p is a more (possibly much more) active enzyme than ATE1-2p.

**Neither ATE1-1p nor ATE1-2p confers metabolic instability on Cys-βgal.** Cysteine is a stabilizing residue in the yeast N-end rule but a secondary destabilizing residue in multicellular organisms such as mammals and amphibians (14, 18, 30, 54). The presence of two alternative regions in the two forms of mouse R-transferase (Fig. 4C) initially suggested that one R-transferase might be specific for N-terminal Asp and Glu, with the other specific for Cys. However, Cys-βgal, which is long-lived in wild-type *S. cerevisiae* (5), remained long-lived in the presence of either ATE1-1p or ATE1-2p (Fig. 6A). This finding and, more directly, the results of amino acid sequencing (see below) suggest the existence of a mammalian tRNA-dependent enzyme (presumably a distinct R-transferase) (18) that mediates destabilizing activity of N-terminal Cys.

**Mouse ATE1-1p and ATE1-2p destabilize Asp-βgal and Glu-βgal through arginylation of their N-terminal residues.** To verify directly that mouse ATE1-1p and ATE1-2p in fact possess the R-transferase activity, we constructed the *ate1Δ ubr1Δ S. cerevisiae* double mutant AVY34, which lacked both R-transferase and N-recognin (E3), the main recognition component of the N-end rule pathway (see Materials and Methods). Consequently, N-terminal arginylation of a test protein in this mutant by an exogenous R-transferase would not result in degradation of the protein, thereby making it possible to isolate enough of the test protein for N-terminal sequencing. Strain AVY34 was transformed with pUB23-D (expressing Ub-Asp-βgal) and also with either pAT1 (expressing ATE1-1p), pAT2 (expressing ATE1-2p), or vector alone and was grown in SG medium. Asp-βgal proteins isolated from these transformants were subjected to N-terminal sequencing (see Materials and Methods). The results (Table 1) directly confirmed that both ATE1-1p and ATE1-2p possessed R-transferase activity. In agreement with the finding that ATE1-1p was more active than ATE1-2p in destabilizing Asp-βgal in vivo (Fig. 6A), Asp-βgal from cells expressing ATE1-1p was

TABLE 1. N-terminal sequencing of X-βgal proteins isolated from *ate1Δ ubr1Δ S. cerevisiae* expressing different R-transferases

| Substrate | Coexpressed protein | N-terminal sequence | Yield (%) |
|---|---|---|---|
| D-$c^K$-βgal | | D-H-G-S-A- | |
| D-$c^K$-βgal | Vector alone | D-H-G-S-A- | ~100 |
| D-$c^K$-βgal | Mouse ATE1-1p | R-D-H-G-S-A- | ~100 |
| D-$c^K$-βgal | Mouse ATE1-2p | R-D-H-G-S-A- | ~50 |
| | | D-H-G-S-A- | ~50 |
| C-$c^K$-βgal | | C-H-G-S-A- | |
| C-$c^K$-βgal | Vector alone | C-H-G-S-A- | ~20 |
| C-$c^K$-βgal | Mouse ATE1-1p | C-H-G-S-A- | ~20 |
| C-$c^K$-βgal | Mouse ATE1-2p | C-H-G-S-A- | ~20 |

found to be completely arginylated, whereas Asp-βgal from cells expressing ATE1-2p was arginylated to approximately 50% (Table 1).

We also determined, using the above procedure, whether mouse ATE1-1p or ATE1-2p could arginylate N-terminal Cys. Approximately 80% of Cys-βgal isolated from *ate1Δ ubr1Δ S. cerevisiae* was found to be N-terminally blocked, presumably acetylated (Table 1). However, the rest of Cys-βgal (~20%) bore the N-terminal sequence beginning with Cys and lacking N-terminal Arg, in agreement with the results of the in vivo Cys-βgal degradation assays (Fig. 6 and Table 1). Thus, both ATE1-1p and Ate1-2p are apparently unable to utilize N-terminal Cys as a substrate in *S. cerevisiae*.

**ATE1-2p is exclusively cytosolic, whereas ATE1-1p is present in either the nucleus or the cytosol.** To determine the intracellular location of the two forms of mouse R-transferase, we constructed fusions to the N terminus of GFP and transiently expressed them in NIH 3T3 cells. Whereas the free 26-kDa GFP was located in both the nucleus and the cytosol (data not shown), the 85-kDa Ate1-2p–GFP fusion was exclusively cytosolic in all of the many transfected cells examined (Fig. 7a to c). In contrast, the 85-kDa ATE1-1p–GFP (the alternative form of R-transferase that is much more active enzymatically than ATE1-2p) was found to be localized differently in different cells on the same coverslip, possibly depending on their cell cycle position and/or metabolic state. Specifically, in ~50% of the transfected cells, ATE1-1p–GFP was exclusively cytosolic (Fig. 7d and e), as was ATE1-2p–GFP (Fig. 7a to c), but in the other ~50% of cells, ATE1-1p–GFP was present in the nucleus as well and, moreover, appeared to be significantly enriched in the nucleus (Fig. 7f and g). Thus, the two 43-residue alternative regions in ATE1-1p and ATE1-2p confer overlapping but nonidentical intracellular distributions on the respective R-transferases.

While the nonuniformity of the ATE1-1p–GFP localization among mouse cells in a single culture remains to be understood, its preferential location in the nuclei of some cells is consistent with a high content of basic residues in its 43-residue region, in comparison to the alternative homologous region of ATE1-2p (Fig. 2C). (No sequences fitting the consensus sequences of known nuclear localization signals could be detected in the 43-residue region of ATE1-1p). In contrast to mouse R-transferases, *S. cerevisiae* Ate1p was shown to be located predominantly in the nuclei of yeast cells (56a).

**The ratio of *Ate1-1* to *Ate1-2* mRNA varies greatly among mouse tissues.** Northern hybridization, using the 1.1-kb mouse *Ate1* cDNA fragment (nt 638 to 1734) as a probe, detected a

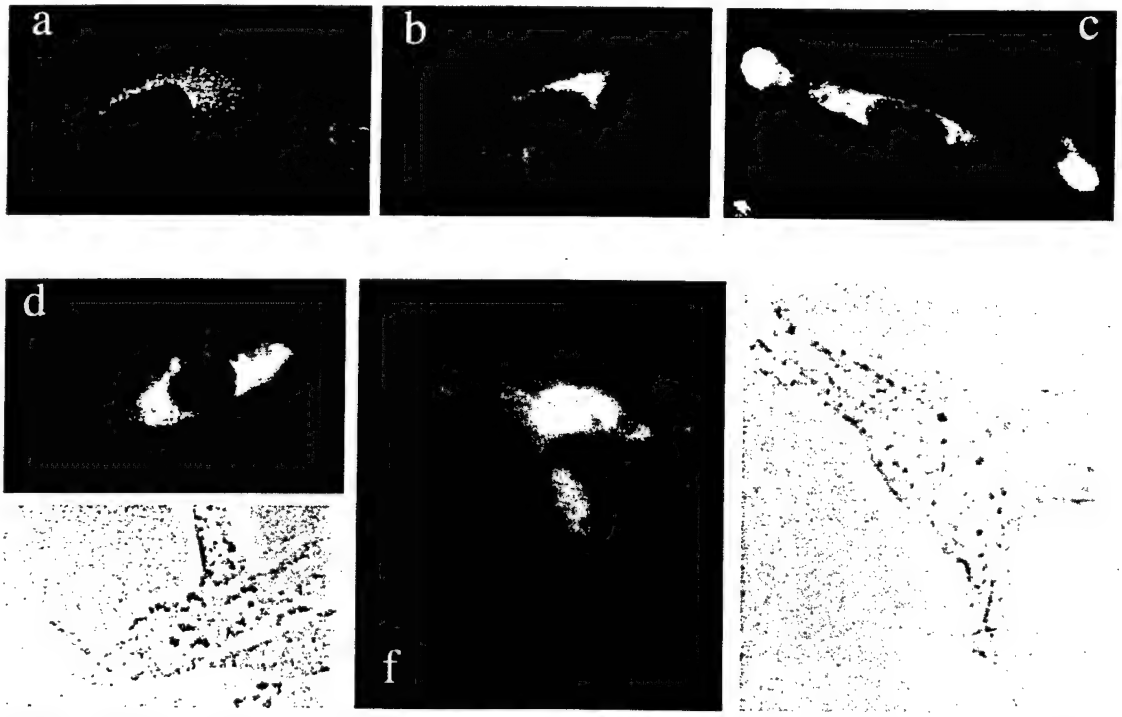



FIG. 7. Intracellular localization of mouse ATE1-1p and ATE1-2p. Shown are green (GFP) fluorescence (a to d and f) and phase-contrast (e and g) micrographs of mouse NIH 3T3 cells transiently transfected with ATE1-1p-GFP (d to g) or ATE1-2p–GFP (a to c) fusion proteins (see Materials and Methods). Panels a to c show different examples of the exclusively cytosolic localization of ATE1-2p. Regions around the nucleus and in the lamellar protrusions at the edges of a cell (c) exhibit higher GFP fluorescence, possibly because of a greater thickness of cells in these areas. Panels d plus e and f plus g show pairs of GFP fluorescence and phase-contrast pictures of cells that express ATE1-1p-GFP. The cell in panels d and e shows ATE1-1p-GFP in the cytosol but not in the nucleus. Cells in panels f and g contain ATE1-1p-GFP in both the cytosol and the nucleus, the latter being apparently enriched in ATE1-1p-GFP.

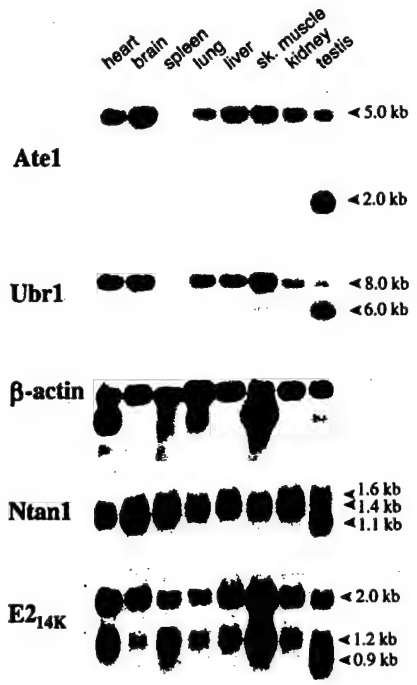



FIG. 8. Northern hybridization analyses of mouse *Ate1* mRNA. The Northern blots of mRNA from different mouse tissues were probed with an *Ate1* cDNA fragment (nt 638 to 1734) which can hybridize to both forms of *Ate1* mRNA. A mouse β-actin cDNA probe was used for comparing the total RNA loads as described previously (19). The same blot was also hybridized with mouse *Ubr1* (29) (see the introduction). The apparent absence of *Ate1* and *Ubr1* mRNAs from the spleen is an artifact of RNA degradation in this lane of the blot (data not shown). Also shown are the results of analogous Northern hybridizations of the mouse *Ntan1* cDNA, encoding the Asn-specific Nt-amidase (Fig. 1A), and $E2_{14K}$, encoding the relevant Ub-conjugating (E2) enzyme (19). The approximate sizes of transcripts are indicated on the right.

single ~5.0-kb transcript (a mixture of the *Ate1-1* and *Ate1-2* mRNAs) in all of the mouse tissues examined except the testis, where the ~5-kb *Ate1* mRNA was a minor one, the major species being ~2 kb (Fig. 8). Both ATE1p and the other targeting components of the mammalian N-end rule pathway are expressed ubiquitously (at various levels), and the testis-specific patterns of transcripts are characteristic for all of them as well (Fig. 8). The existence of the Y-chromosome-encoded, testis-specific variant of the Ub-activating (E1) enzyme (27, 35) suggests that the testis-specific modifications of the N-end rule pathway may be functionally relevant in spermatogenesis.

To determine the ratio of *Ate1-1* to *Ate1-2* mRNA in different mouse tissues or cells in culture, we employed RT-PCR, using sequence differences between the two alternative, homologous 129-bp exons to distinguish between them (see Materials and Methods) (Fig. 3). Approximately equal amounts of *Ate1-1* and *Ate1-2* mRNAs were present in mouse embryonic fibroblasts and in human 293 cells in culture (Fig. 3 and data not shown). However, the molar ratio of *Ate1-1* to *Ate1-2* mRNA was found to vary greatly among the mouse tissues: it was ~0.1 in the skeletal muscle, ~0.25 in the spleen, ~3.3 in the liver, and brain, and ~10 in the testis (Fig. 3). Thus, while the total expression of *Ate1* (*Ate1-1* plus *Ate1-2*) varies by 2- to 4-fold among mouse tissues (Fig. 8), the difference in expression levels between *Ate1-1* and *Ate1-2* mRNAs can be as high as a 100-fold (the skeletal muscle versus the testis) (Fig. 3), suggesting that the two forms of R-transferase may be functionally distinct.

## DISCUSSION

The N-end rule pathway is one of several proteolytic pathways of the Ub system (23, 54, 55). Among the targets of the N-end rule pathway are proteins that bear destabilizing N-terminal residues. In the yeast *S. cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by a specific amidase (6), to the secondary destabilizing N-terminal residues Asp and Glu. The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *ATE1*-encoded R-transferase, to Arg, one of the primary destabilizing residues (7) (Fig. 1B). In mammals, the set of secondary destabilizing residues contains not only Asp and Glu but also Cys, which is a stabilizing residue in yeast (18, 54) (Fig. 1).

In this work, we isolated cDNA encoding the mouse R-transferase, ATE1p, and found that this enzyme exists in two forms, termed ATE1-1p and ATE1-2p, which differ by containing one of the two alternative, homologous 43-residue regions. The two 516-residue R-transferases are produced from the mouse *Ate1* gene by a pathway of alternative splicing that retains one or the other of the two homologous 129-bp exons. The presence of two adjacent, homologous, equal-length, and alternatively utilized exons in a gene (Fig. 4) is nearly unprecedented. To our knowledge, just one such case was described previously: the mouse κE2 enhancer-binding protein E12/E47 (37). The two κE2-binding proteins, E12 and E47, are produced through a switch between two alternative, equal-length exons, resulting in two helix-loop-helix DNA-binding proteins that differ in the ability to homodimerize. Specifically, E47 can bind to the κE2 enhancer either as a homodimer or as a heterodimer with MyoD, whereas E12 can bind as a heterodimer with MyoD but not as a homodimer (37).

We report the following major findings.

(i) Identification, through species walking, and isolation of the mouse cDNA encoding R-transferase (or ATE1p) have shown that mammalian ATE1p exists in two forms, ATE1-1p and ATE1-2p, which differ exclusively by one of the two alternative, homologous 43-residue regions (Fig. 2A).

(ii) The corresponding alternative 129-bp exons are adjacent in the mouse *Ate1* gene. Moreover, the 12-bp sequences around the splice acceptor sites of these exons (6 bp in the intron and 6 bp in the exon) are identical between the two exons (Fig. 4). The splicing of *Ate1* pre-mRNA proceeds in such way that one, and only one, of the alternative 129-bp exons is always retained in the mature *Ate1* mRNA.

(iii) The human *ATE1* gene also contains the two alternative 129-bp exons, whereas the plant (*A. thaliana*) and fly (*D. melanogaster*) *Ate1* genes encode a single form of ATE1p (Fig. 2 and 5). The corresponding 43-residue regions are significantly similar among all of the sequenced R-transferases, from *S. cerevisiae* to mammals (Fig. 2C). The set of *Ate1* genes from mammals to yeast defines a distinct family of proteins, the ATE family. The splicing-derived alternative forms of R-transferase have evolved apparently after the divergence of the arthropod and vertebrate lineages.

(iv) Expression of the mouse *Ate1-1* and *Ate1-2* cDNAs in *ate1Δ S. cerevisiae*, and N-terminal sequencing of isolated X-βgal test proteins, was used to show that ATE1-1p and ATE1-2p could implement the Asp/Glu-specific subset of the N-end rule pathway and that they did so through the arginylation of N-terminal Asp or Glu in the test substrates (Fig. 6A and Table 1).

(v) While the destabilizing activity of the mouse ATE1-1p R-transferase is only slightly lower than that of *S. cerevisiae* R-transferase, the activity of mouse ATE1-2p is significantly

(possibly considerably) lower than that of ATE1-1p. This conclusion follows also from a comparison of the N-terminal arginylation of Asp-βgal by the two R-transferases (Table 1). The results of coexpressing mouse ATE1-1p and ATE1-2p in the same *ate1*Δ yeast cells were consistent with the conjecture that R-transferase functions as a monomer (Fig. 6B).

(vi) Neither ATE1-1p nor ATE1-2p could confer instability on (or arginylate) Cys-βgal in *ate1*Δ *S. cerevisiae* (Fig. 6A and Table 1). Cys is a stabilizing residue in yeast but a secondary destabilizing residue in the mammalian N-end rule (54). A distinct Cys-specific mammalian R-transferase suggested by these data remains to be identified.

(vii) Mouse ATE1-2p (tested as a GFP fusion) was exclusively cytosolic in mouse 3T3 cells, whereas ATE1-1p was localized differentially in different cells of the same (unsynchronized) culture: it was either exclusively cytosolic or present in both the cytosol and the nucleus (Fig. 7).

(viii) Mouse *Ate1* is a ubiquitously expressed gene. A single ~5-kb mRNA was present in all of the tissues examined except the testis, where the major *Ate1* transcript was ~2 kb in length (Fig. 8). The testis-specific differential expression patterns are also characteristic of the other targeting components of the mammalian N-end rule pathway, such as the *Ntan1*-encoded Asn-specific Nt-amidase and the *Ubr1*-encoded N-recognin (E3α) (19, 29).

(ix) The molar ratio of *Ate1-1* to *Ate1-2* mRNA varies up to a 100-fold among different mouse tissues (Fig. 3 and Results), suggesting a functional significance of the difference between the two R-transferases.

The region of ATE1p that corresponds to the two 129-bp mammalian *Ate1* exons has been significantly conserved throughout eukaryotic evolution, Tyr-296, Gln-297, and His-301 of the mouse ATE1-1p being among the most highly conserved residues (Fig. 2C). No putative members of the ATE family could be detected among the currently known prokaryotic ORFs. The most highly conserved region of R-transferases is an 82-residue stretch (residues 336 to 417) of mouse ATE1p: this region is 95, 76, and 63% identical to the corresponding regions of the human, *D. melanogaster*, and *A. thaliana* ATE1p, respectively (Fig. 2B). A Cys residue(s) is likely to be a component of the active site of R-transferase (31, 32). Among the five fully conserved Cys residues in proteins of the ATE family, four are located in the 56-residue N-terminal region (residues 23 to 78 of mouse ATE1p) (Fig. 2B). Conversion of some of these cysteines in *S. cerevisiae* Ate1p to alanines was found to decrease greatly the R-transferase activity of yeast Ate1p (16a, 32). Furthermore, derivatives of mouse ATE1-1p and ATE1-2p that lacked the first 42 residues were completely inactive in the yeast-based Asp-βgal degradation assay of a kind described in Fig. 7 (data not shown). Finally, a 90-residue C-terminal truncation of *S. cerevisiae* Ate1p did not result in a major decrease of its R-transferase activity (16a). Thus, the active site of R-transferase is likely to encompass at least some of the above N-terminal cysteines.

Since the two mammalian R-transferases (Fig. 4C) are identical in size and, except for a 43-residue region, are identical otherwise as well, it is likely that the previously described (partially purified) mammalian R-transferases (12, 47) were in fact mixtures of Ate1-1p and Ate1-2p. On the other hand, fractionation of a crude R-transferase preparation from rabbit reticulocytes did yield, in addition to a major fraction of R-transferase, chromatographically distinct R-transferase fractions that were not investigated further (12). The ratio of Ate1-1p to Ate1-2p in rabbit reticulocytes is currently unknown.

A splicing-mediated switch that replaces the 129-bp *Ate1* exon of ATE1-1p with the alternative 129-bp exon results in a protein, ATE1-2p, that has a significantly (possibly considerably) lower R-transferase activity (Fig. 6). In addition, ATE1-2p is unable to enter the nucleus (as a GFP fusion), in contrast to ATE1-1p (Fig. 7). Taken together with the finding that the expression ratio of the two *Ate1* mRNAs, *Ate1-1* and *Ate1-2*, varies up to a 100-fold among different mouse tissues (Fig. 3 and Results), these data suggest that the two R-transferases are functionally distinct as well. Cited below are some of the possibilities that are consistent with the available evidence.

Mouse ATE1-2p has the R-transferase activity but arginylates, at steady state, only ~50% of Asp-βgal in *ate1*Δ *S. cerevisiae*, in contrast to both ATE1-2p and *S. cerevisiae* Ate1p (Fig. 6 and Table 1). Moreover, the inefficient arginylation by ATE1-2p occurs in spite of its overexpression in *S. cerevisiae*. In contrast, the yeast Ate1p, which in wild-type *S. cerevisiae* is a weakly expressed protein (7), can quantitatively arginylate in vivo an overexpressed substrate such as Asp-βgal (18). Thus, at a low level of expression (which is likely to be the case in the mouse), the ATE1-2p R-transferase may be, in effect, an inactive enzyme, in contrast to ATE1-1p. If so, ATE1-2p might act as an (indirect) inhibitor of the ATE1-1p function, for example, through a competition with ATE1-1p for the binding to a component of the targeting complex in the N-end rule pathway. (The apparent absence of such competition in *S. cerevisiae* [Fig. 6B] may result from the lack of binding by ATE1p to heterologous yeast proteins.) It is also possible that a large difference in activity between mouse ATE1-1p and ATE1-2p in yeast reflects not their different enzymatic activities in the mouse but a (physiologically irrelevant) differential recognition of an essential yeast cofactor such as Arg-tRNA. Direct comparisons of arginylation kinetics by the purified mouse and yeast R-transferases will be required to address this unlikely but unexcluded interpretation.

Another possibility is that ATE1-2p has a distinct enzymatic activity that has been missed by the current N-terminal arginylation assay (Fig. 6 and Table 1). For example, ATE1-2p might be able to arginylate an internal residue in a substrate protein. In vitro enzymological dissection of ATE1-1p and ATE1-2p will address this and related conjectures. Yet another possibility is that the alternative 43-residue regions of ATE1-1p and ATE1-2p confer different metabolic stabilities on the two R-transferases, the lower apparent activity of ATE1-2p in yeast being due at least in part to its shorter half-life. A test of this model in mouse cells requires antibodies specific for the alternative regions of the two R-transferases; preparation of such antibodies is under way.

In yeast, the N-end rule pathway is present in both the cytosol and the nucleus. The apparent exclusion of mouse ATE1-2p from the nucleus and the different ratios of *Ate1-1* to *Ate1-2* mRNA among the mouse tissues suggest that the rule book of the N-end rule pathway may be regulated differentially in the cytosol and the nucleus, through a cell-type-specific expression of the pathway's components that are located in one but not the other compartment.

Physiological substrates for either eukaryotic R-transferases (54) or their prokaryotic counterparts, L, F-transferases (1, 25, 45) are not known. The cloning and characterization of the first mammalian *Ate1* cDNAs and genes (Fig. 2), and the discovery of alternative splicing that yields mouse ATE1-1p and ATE1-2p (Fig. 4) should facilitate understanding of the functions of mammalian R-transferases, in part through the analysis of ATE1-1p and ATE1-2p enzymes and also because it is now possible to construct mouse strains that lack ATE1-1p and/or ATE1-2p.

## ACKNOWLEDGMENTS

The first two authors contributed equally to this work.

We thank Gary Hathaway of the Caltech Microchemistry Facility for the sequencing of X-βgal proteins. We are grateful to Hai Rao, Glenn Turner, Fangyong Du, and Lawrence Peck for helpful suggestions and to Fangyong Du, Federico Navarro-Garcia, Hai Rao, and Youming Xie for comments on the manuscript.

This work was supported by grants DK39520 and GM31530 to A.V. from the National Institutes of Health.

## REFERENCES

1. **Abramochkin, G., and T. E. Shrader.** 1995. The leucyl/phenylalanyl-tRNA-protein transferase. Overexpression and characterization of substrate recognition, domain structure, and secondary structure. J. Biol. Chem. **270:**20621–20628.
2. **Alagramam, K., F. Naider, and J. M. Becker.** 1995. A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. Mol. Microbiol. **15:**225–234.
3. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl (ed.).** 1996. Current protocols in molecular biology. Wiley-Interscience, New York, N.Y.
4. **Bachmair, A., D. Finley, and A. Varshavsky.** 1986. *In vivo* half-life of a protein is a function of its amino-terminal residue. Science **234:**179–186.
5. **Bachmair, A., and A. Varshavsky.** 1989. The degradation signal in a short-lived protein. Cell **56:**1019–1032.
6. **Baker, R. T., and A. Varshavsky.** 1995. Yeast N-terminal amidase: a new enzyme and component of the N-end rule pathway. J. Biol. Chem. **270:**12065–12074.
7. **Balzi, E., M. Choder, W. Chen, A. Varshavsky, and A. Goffeau.** 1990. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. J. Biol. Chem. **265:**7464–7471.
8. **Bartel, B., I. Wünning, and A. Varshavsky.** 1990. The recognition component of the N-end rule pathway. EMBO J. **9:**3179–3189.
9. **Baumeister, W., J. Walz, F. Zühl, and E. Seemüller.** 1998. The proteasome: paradigm of a self-compartmentalizing protease. Cell **92:**367–380.
10. **Byrd, C., G. C. Turner, and A. Varshavsky.** 1998. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. EMBO J. **17:**269–277.
11. **Chau, V., J. W. Tobias, A. Bachmair, D. Marriott, D. J. Ecker, D. K. Gonda, and A. Varshavsky.** 1989. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science **243:**1576–1583.
12. **Ciechanover, A., S. Ferber, D. Ganoth, S. Elias, A. Hershko, and S. Arfin.** 1988. Purification and characterization of arginyl-tRNA-protein transferase from rabbit reticulocytes. J. Biol. Chem. **263:**11155–11167.
13. **Coux, O., K. Tanaka, and A. L. Goldberg.** 1996. Structure and functions of the 20S and 26S proteasomes. Annu. Rev. Biochem. **65:**801–817.
14. **Davydov, I. V., D. Patra, and A. Varshavsky.** The N-end rule pathway in *Xenopus* oocyte extracts. Arch. Biochem. Biophys., in press.
15. **Dayal, V. K., G. Chakraborty, J. A. Sturman, and N. A. Ingoglia.** 1990. The site of amino acid addition to posttranslationally modified proteins in regenerating rat sciatic nerves. Biochim. Biophys. Acta **1038:**172–177.
16. **deGroot, R. J., T. Rümenapf, R. J. Kuhn, and J. H. Strauss.** 1991. Sindbis virus RNA polymerase is degraded by the N-end rule pathway. Proc. Natl. Acad. Sci. USA **88:**8967–8971.

16a. **Du, F., and A. Varshavsky.** Unpublished data.

17. **Ferber, S., and A. Ciechanover.** 1987. Role of arginine-tRNA in protein degradation by the ubiquitin pathway. Nature **326:**808–811.
18. **Gonda, D. K., A. Bachmair, I. Wünning, J. W. Tobias, W. S. Lane, and A. Varshavsky.** 1989. Universality and structure of the N-end rule. J. Biol. Chem. **264:**16700–16712.
19. **Grigoryev, S., A. E. Stewart, Y. T. Kwon, S. M. Arfin, R. A. Bradshaw, N. A. Jenkins, N. J. Copeland, and A. Varshavsky.** 1996. A mouse amidase specific for N-terminal asparagine: the gene, the enzyme, and their function in the N-end rule pathway. J. Biol. Chem. **271:**28521–28532.

19a. **Grigoryev, S., and A. Varshavsky.** Unpublished data.

20. **Haas, A. J., and T. J. Siepman.** 1997. Pathways of ubiquitin conjugation. FASEB J. **11:**1257–1268.
21. **Hershko, A.** 1991. The ubiquitin pathway for protein degradation. Trends Biochem. Sci. **16:**265–268.
22. **Hill, C. P., N. L. Johnston, and R. E. Cohen.** 1993. Crystal structure of a ubiquitin-dependent degradation substrate: a three-disulfide form of lysozyme. Proc. Natl. Acad. Sci. USA **90:**4136–4140.
23. **Hochstrasser, M.** 1996. Ubiquitin-dependent protein degradation. Annu. Rev. Genet. **30:**405–439.
24. **Hondermarck, H., J. Sy, R. A. Bradshaw, and S. M. Arfin.** 1992. Dipeptide inhibitors of ubiquitin-mediated protein turnover prevent growth factor-induced neurite outgrowth in rat pheochromocytoma PC12 cells. Biochem. Biophys. Res. Commun. **30:**280–288.
25. **Ichetovkin, I. L., G. Abramochkin, and T. E. Shrader.** 1997. Substrate recognition by the leucyl/phenylalanyl-tRNA protein transferase: conservation within the enzyme family and localization to the trypsin-resistant domain. J. Biol. Chem. **272:**33009–33014.

25a. **Johnston, J., and A. Varshavsky.** Unpublished data.

26. **Kaiji, H., G. D. Novelli, and A. Kaiji.** 1963. A soluble amino acid-incorporating system from rat liver. Biochim. Biophys. Acta **76:**474–479.
27. **Kay, G. F., A. Ashworth, G. D. Penny, M. Dunlop, S. Swift, N. Brockdorff, and S. Rastan.** 1991. A candidate spermatogenesis gene on the mouse Y chromosome is homologous to ubiquitin-activating enzyme E1. Nature **354:**486–489.
28. **King, R. W., R. J. Deshaies, J. M. Peters, and M. W. Kirschner.** 1996. How proteolysis drives the cell cycle. Science **274:**1652–1659.

28a. **Kwon, Y. T., V. Denenberg, and A. Varshavsky.** Unpublished data.

29. **Kwon, Y. T., Y. Reiss, V. A. Fried, A. Hershko, J. K. Yoon, D. K. Gonda, P. Sangan, N. G. Copeland, N. A. Jenkins, and A. Varshavsky.** 1998. The mouse and human genes encoding the recognition component of the N-end rule pathway. Proc. Natl. Acad. Sci. USA **95:**7898–7903.

29a. **Kwon, Y. T., and A. Varshavsky.** Unpublished data.

30. **Levy, F., N. Johnsson, T. Rümenapf, and A. Varshavsky.** 1996. Using ubiquitin to follow the metabolic fate of a protein. Proc. Natl. Acad. Sci. USA **93:**4907–4912.
31. **Li, J., and C. Pickart.** 1995. Inactivation of arginyl-tRNA protein transferase by a bifunctional arsenoxide: identification of residues proximal to arsenoxide site. Biochemistry **34:**139–147.
32. **Li, J., and C. M. Pickart.** 1995. Binding of phenylarsenoxide to Arg-tRNA-protein transferase is independent of vicinal thiols. Biochemistry **34:**15829–15837.
33. **Madura, K., R. J. Dohmen, and A. Varshavsky.** 1993. N-recognin/Ubc2 interactions in the N-end rule pathway. J. Biol. Chem. **268:**12046–12054.
34. **Madura, K., and A. Varshavsky.** 1994. Degradation of Gα by the N-end rule pathway. Science **265:**1454–1458.
35. **Mitchell, M. J., D. R. Woods, P. K. Tucker, J. S. Opp, and C. E. Bishop.** 1991. Homology of a candidate spermatogenesis gene from the mouse Y chromosome to the ubiquitin-activating enzyme E1. Nature **354:**483–486.
36. **Mumberg, G., R. Müller, and M. Funk.** 1994. Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res. **22:**5767–5768.
37. **Murre, C., P. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell **56:**777–783.
38. **Ota, I. M., and A. Varshavsky.** 1993. A yeast protein similar to bacterial two-component regulators. Science **262:**566–569.
39. **Pickart, C. M.** 1997. Targeting of substrates to the 26S proteasome. FASEB J. **11:**1055–1066.

39a. **Rao, H., and A. Varshavsky.** Unpublished data.

40. **Rechsteiner, M., L. Hoffman, and W. Dubiel.** 1993. The multicatalytic and 26S proteases. J. Biol. Chem. **268:**6065–6068.
41. **Reiss, Y., and A. Hershko.** 1990. Affinity purification of ubiquitin-protein ligase on immobilized protein substrates. J. Biol. Chem. **265:**3685–3690.
42. **Rothstein, R.** 1991. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol. **194:**281–301.
43. **Rubin, D. M., S. van Nocker, M. Glickman, O. Coux, I. Wefes, S. Sadis, H. Fu, A. Goldberg, R. Vierstra, and D. Finley.** 1997. ATPase and ubiquitin-binding proteins of the yeast proteasome. Mol. Biol. Rep. **24:**17–26.
44. **Scheffner, M., S. Smith, and S. Jentsch.** 1998. The ubiquitin conjugation system, p. 65–98. *In* J.-M. Peters, J. R. Harris, and D. Finley (ed.), Ubiquitin and the biology of the cell. Plenum Press, New York, N.Y.
45. **Shrader, T. E., J. W. Tobias, and A. Varshavsky.** 1993. The N-end rule in *Escherichia coli*: cloning and analysis of the leucyl, phenylalanyl-tRNA-protein transferase gene *aat*. J. Bacteriol. **175:**4364–4374.
46. **Sijst, A. J. A. M., I. Pilip, and E. G. Pamer.** 1997. The *Listeria monocytogenes*-secreted p60 protein is an N-end rule substrate in the cytosol of infected cells. J. Biol. Chem. **272:**19261–19268.
47. **Soffer, R. L.** 1980. Biochemistry and biology of aminoacyl-tRNA-protein transferases, p. 493–505. *In* D. Söll, J. Abelson, and P. R. Schimmel (ed.), Transfer RNA: biological aspects. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
48. **Solomon, V., V. Baracos, P. Sarraf, and A. Goldberg.** When muscles atrophy, rates of ubiquitin conjugation increase, largely through activation of the N-end rule pathway. Proc. Natl. Acad. Sci. USA, in press.
49. **Solomon, V., S. H. Lecker, and A. L. Goldberg.** 1998. The N-end rule pathway mediates a major fraction of protein degradation in skeletal muscle. J. Biol. Chem. **273:**25216–2522.
50. **Stewart, A.** 1995. Trends in genetics nomenclature guide. Elsevier Science, Ltd., Cambridge, United Kingdom.
51. **Stewart, A. E., S. M. Arfin, and R. A. Bradshaw.** 1995. The sequence of porcine protein N-terminal asparagine amidohydrolase: a new component of the N-end rule pathway. J. Biol. Chem. **270:**25–28.
52. **Taban, C. H., H. Hondermarck, R. A. Bradshaw, and B. Boilly.** 1996. Effect of a dipeptide inhibiting ubiquitin-mediated protein degradation on nerve-dependent limb regeneration in the newt. Experientia **52:**865–870.
53. **Tobias, J. W., T. E. Shrader, G. Rocap, and A. Varshavsky.** 1991. The N-end rule in bacteria. Science **254:**1374–1377.

54. **Varshavsky, A.** 1997. The N-end rule pathway of protein degradation. Genes Cells **2**:13–28.
55. **Varshavsky, A.** 1997. The ubiquitin system. Trends Biochem. Sci. **22**:383–387.
56. **Varshavsky, A., C. Byrd, I. V. Davydov, R. J. Dohmen, F. Du, M. Ghislain, M. Gonzalez, S. Grigoryev, E. S. Johnson, N. Johnsson, J. A. Johnston, Y. T. Kwon, F. Lévy, O. Lomovskaya, K. Madura, I. Ota, T. Rümenapf, T. E. Shrader, T. Suzuki, G. Turner, P. R. H. Waller, and A. Webster.** 1998. The N-end rule pathway, p. 223–278. *In* J.-M. Peters, J. R. Harris, and D. Finley (ed.), Ubiquitin and the biology of the cell. Plenum Press, New York, N.Y.
56a. **Wang, H. R., and A. Varshavsky.** Unpublished data.
57. **Wang, Y. M., and N. A. Ingoglia.** 1997. N-terminal arginylation of sciatic nerve and brain proteins following injury. Neurochem. Res. **22**:1453–1459.

Molecular Biology of the Cell
Vol. 10, 329–344, February 1999

# Detection of Transient In Vivo Interactions between Substrate and Transporter during Protein Translocation into the Endoplasmic Reticulum

**Martin Dünnwald,* Alexander Varshavsky,† and Nils Johnsson*‡**

*Max-Delbrück-Laboratorium, D-50829 Köln, Germany; and †Division of Biology, California Institute of Technology, Pasadena, California 91125

Submitted October 2, 1998; Accepted November 11, 1998
Monitoring Editor: Peter Walter

The split-ubiquitin technique was used to detect transient protein interactions in living cells. $N_{ub}$, the N-terminal half of ubiquitin (Ub), was fused to Sec62p, a component of the protein translocation machinery in the endoplasmic reticulum of *Saccharomyces cerevisiae*. $C_{ub}$, the C-terminal half of Ub, was fused to the C terminus of a signal sequence. The reconstitution of a quasi-native Ub structure from the two halves of Ub, and the resulting cleavage by Ub-specific proteases at the C terminus of $C_{ub}$, serve as a gauge of proximity between the two test proteins linked to $N_{ub}$ and $C_{ub}$. Using this assay, we show that Sec62p is spatially close to the signal sequence of the prepro-α-factor in vivo. This proximity is confined to the nascent polypeptide chain immediately following the signal sequence. In addition, the extent of proximity depends on the nature of the signal sequence. $C_{ub}$ fusions that bore the signal sequence of invertase resulted in a much lower Ub reconstitution with $N_{ub}$-Sec62p than otherwise identical test proteins bearing the signal sequence of prepro-α-factor. An inactive derivative of Sec62p failed to interact with signal sequences in this assay. These in vivo findings are consistent with Sec62p being part of a signal sequence-binding complex.

## INTRODUCTION

A critical step during the translocation of a protein across the membrane of the endoplasmic reticulum (ER) is the interaction between the signal sequence of a nascent polypeptide and its receptors (Walter *et al.*, 1981; Gilmore and Blobel, 1985; Walter and Johnson, 1994). A stretch of 8 to 12 hydrophobic residues, often at the N terminus of a protein, comprises a signal sequence that is sufficient to initiate the protein's translocation into the endoplasmic reticulum (ER) (Rapoport *et al.*, 1996). To be compatible with a high flux of polypeptides through a limited number of translocation channels in the ER membrane, the interaction between the signal sequence and its receptors has to be short lived. Its transient nature makes such a receptor–ligand interaction difficult to study, especially in living cells. The approaches used for the analysis of protein translocation in cell-free systems circumvent the transience of the signal sequence–receptor interaction by pausing or stopping the synthesis of a nascent polypeptide chain at different stages of its movement to and across the ER membrane (Krieg *et al.*, 1986; Kurzchalia *et al.*, 1986; Connolly *et al.*, 1989). Given these constraints, it is essential to verify in vivo the models derived from in vitro studies. The ability to analyze early translocation events in vivo should also be important for defining the immediate environment of the nascent chain on its path from the ribosome to the ER membrane.

Most of the current methods for detecting protein interactions in vivo either do not operate at the ER membrane or are unable to detect a transient proximity between proteins (Fields and Song, 1989; Aronheim *et al.*, 1997; Rossi *et al.*, 1997; Miyawaki *et al.*, 1997). In the present work, we show that the previously developed split-ubiquitin (split-Ub) technique, also called USPS (Ub/split/protein/sensor) (Johnsson and Varshavsky, 1994a, 1997), is capable of detecting a transient in vivo

‡ Corresponding author. E-mail address: johnsson@mpiz-koeln.mpg.de.

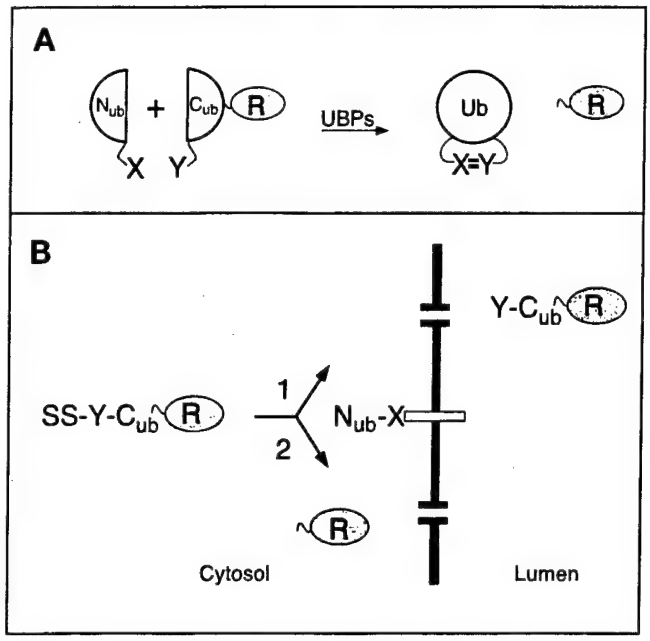



**Figure 1.** The split-Ub technique and its application to the analysis of protein translocation. (A) $N_{ub}$ and $C_{ub}$ are linked to the interacting proteins X and Y. The X–Y complex brings $N_{ub}$ and $C_{ub}$ into close proximity. $N_{ub}$ and $C_{ub}$ reconstitute a quasi-native Ub moiety, which is cleaved by the UBPs, yielding the free reporter R (Johnsson and Varshavsky, 1994a). (B) Using split-Ub to monitor the proximity between a secretory protein and a component of the translocation machinery. A signal sequence-bearing $C_{ub}$ fusion (SS-Y-$C_{ub}$-R) and a $N_{ub}$ fusion ($N_{ub}$-X) are coexpressed in a cell. Pathway 1: when $N_{ub}$ is linked to a protein not involved in the targeting for translocation, the uncleaved (except for the signal sequence SS) Y-$C_{ub}$-R enters the ER. Pathway 2: when $N_{ub}$ is linked to a protein involved in the targeting for translocation the signal sequence of the SS-Y-$C_{ub}$-R brings $N_{ub}$ and $C_{ub}$ into close contact. As a result, some of the SS-Y-$C_{ub}$-R and $N_{ub}$-X molecules interact to form the quasi-native Ub, yielding the free reporter R in the cytosol.

interaction between polypeptides. The split-Ub method is based on the ability of $N_{ub}$ and $C_{ub}$, the N- and C-terminal halves of Ub, to assemble into a quasi-native Ub. Ub-specific proteases (UBPs), which are present in all eukaryotic cells, recognize the reconstituted Ub, but not its halves, and cleave the Ub moiety off a reporter protein that had been linked to the C terminus of $C_{ub}$. The liberation of the reporter serves as a readout indicating the reconstitution of Ub. The assay is designed in a way that prevents efficient association of $N_{ub}$ and $C_{ub}$ by themselves, but allows it if the two Ub halves are separately linked to proteins that interact in vivo (Figure 1A). The split-Ub assay has been shown to detect the in vivo dimerization of a leucine zipper-containing domain of the Gcn4p transcriptional activator, and the in vivo interaction between two subunits of the oligosaccharyltransferase complex (Johnsson and Varshavsky, 1994a; Stagljar *et al.*, 1998).

In the present work, we focus on the interaction between Sec62p of the yeast *Saccharomyces cerevisiae* and proteins bearing two different signal sequences. Extensive evidence indicates that Sec62p is a component of the ER translocation machinery (Deshaies and Schekman, 1989; Rothblatt *et al.*, 1989; Müsch *et al.*, 1992). Sec62p is a part of the tetrameric Sec62/63 complex that also contains Sec71p and Sec72p (Deshaies *et al.*, 1991; Feldheim and Schekman, 1994). Sec62/63p can be isolated as a tetramer, or as a part of a larger assembly, the heptameric Sec complex (Panzner *et al.*, 1995). In addition to the Sec62/63 complex, the heptamer contains the trimer of Sec61p. This trimer (Sec61p, Sss1p, Sbh1p in yeast; Sec61$\alpha$, Sec61$\beta$, Sec61$\gamma$ in mammals) forms the aqueous channel through which a polypeptide chain is translocated across the ER membrane (Simon and Blobel, 1991; Görlich *et al.*, 1992; Crowley *et al.*, 1993, 1994; Mothes *et al.*, 1994; Hanein *et al.*, 1996; Beckmann *et al.*, 1997).

The role of the Sec62/63 tetramer is less well defined. Cross-linking and reconstitution experiments in vitro have shown that Sec62p is close to the nascent polypeptide chain before the initiation of its translocation (Müsch *et al.*, 1992; Lyman and Schekman, 1997; Matlack *et al.*, 1997). One important role of Sec63p is its ability to recruit the Hsp70-type protein Kar2p of the ER lumen to the vicinity of a translocating polypeptide (Brodsky and Schekman, 1993; Lyman and Schekman, 1997). The Sec62/63 complex is essential for the post-translational translocation of proteins in reconstituted vesicle preparations (Panzner *et al.*, 1995). Genetic analysis supports this conclusion, by showing that the tetrameric Sec62/63 complex is involved in the translocation of proteins whose targeting to the ER membrane is not abolished by the loss of the signal recognition particle (SRP) (Ng *et al.*, 1996). However, it is less clear whether the Sec62/63 complex is the receptor for the signal sequences of those proteins.

In the present work, we demonstrate the ability of the split-Ub assay to detect transient protein interactions in living cells. We show that the assay can monitor a close proximity between Sec62p and a segment of the nascent chain of a signal sequence-bearing protein. The apparent extent of this proximity is influenced by the nature of the signal sequence and the position of $C_{ub}$ in the nascent polypeptide. Our analysis yields a crude map of the environment of the nascent chain during its targeting to and translocation across the ER membrane. Taken together, these findings are the first in vivo evidence that Sec62p, a component of the translocation machinery in the endoplasmic reticulum, is a part of a signal sequence-binding complex.

## MATERIALS AND METHODS

### *Construction of Test Proteins*

The $C_{ub}$ fusions 8–13 (Figure 2) were derived from the construct I of Johnsson and Varshavsky (1994a), which encoded Ub-DHFR-ha and

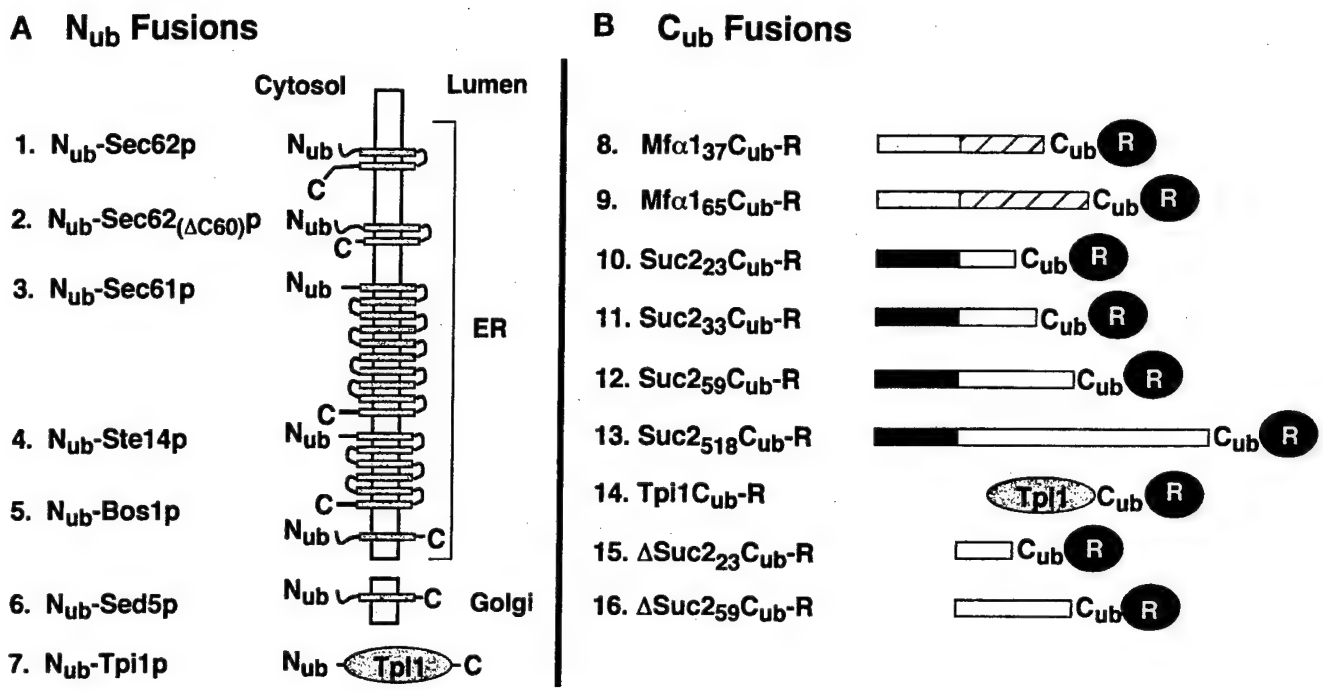



**Figure 2.** $N_{ub}$ and $C_{ub}$ test fusions. (A) $N_{ub}$ (residues 1–36 of Ub) was fused to the N terminus of either a transmembrane protein (constructs 1–6) or a cytosolic protein (construct 7). The N termini of all proteins are located in the cytosol. The orientations and the numbers of membrane-spanning regions (shaded boxes) were derived from the published studies of these proteins, except for Ste14p, for which the exact number of the domains and the localization of the C terminus are not yet known. The $N_{ub}$ fusions 1–5 are located in the membrane of the ER; the $N_{ub}$ fusion 6 resides in the membrane of the early Golgi. The $N_{ub}$ fusion 2 is a Sec62p derivative lacking the C-terminal 60 residues. The $N_{ub}$ fusion 7 contains the full-length triosephosphate isomerase ($N_{ub}$-Tpi1p). (B) $C_{ub}$ fusions. The $C_{ub}$ fusions 8 and 9 contain the signal sequence of the prepro-α-factor (shaded boxes), followed by either 37 (construct 8) or 65 residues (construct 9) of the mature α-factor sequence (striped boxes) and a 7-residue linker sequence (not shown). $C_{ub}$ fusions 10–13 contain the signal sequence of the Suc2p invertase (dark boxes) followed by 23 (construct 10), 33 (construct 11), 59 (construct 12), or 518 residues (construct 13) of the mature sequence of invertase (open boxes) and a 7-residue linker sequence (not shown). The $C_{ub}$ fusion 14 contains the complete sequence of *S. cerevisiae* triosephosphate isomerase (Tpi1p) followed by a 17-residue linker peptide and $C_{ub}$. The $C_{ub}$ fusions 15 and 16 are the signal sequence-lacking counterparts of the fusions 10 and 12. $C_{ub}$ is always followed by a reporter protein. The reporter is DHFR-ha or Ura3p for the $C_{ub}$ fusions 8–13, and DHFR-ha for the $C_{ub}$ fusions 14, 15, and 16.

contained a *Bam*HI site at the amino acid position 36 of Ub, and from the previously described Ub translocation constructs I, VI, IX, X, XXIII, and XXV (Johnsson and Varshavsky, 1994b). The above *Bam*HI site of the Ub-DHFR-ha construct I was fused to a linker sequence in which a 5′-*Sal*I site allowed the in-frame insertion of an *Eag*I–*Sal*I fragment containing the promoter, the signal sequence, and a portion of the mature sequence of the corresponding Ub fusions. The newly introduced sequence was G TCG ACC ATG TCG GGG GGG ATC CCT. The last three triplets encode residues 35, 36, and 37 of Ub (the beginning of $C_{ub}$). The underlined sequences are the *Sal*I and *Bam*HI sites, respectively. The final constructs were in the single-copy plasmids pRS314 or pRS315 (Sikorski and Hieter, 1989). Expression of the $C_{ub}$ fusions bearing Dha (DHFR-ha) as a reporter was mediated by the $P_{ADH1}$ promoter, except for the $C_{ub}$ fusion 14, which was expressed from the $P_{CUP1}$ promoter. The same promoter was used for expressing the Ura3p-based $C_{ub}$ fusions.

The $C_{ub}$ fusions 15 and 16 (Figure 2) were derived from constructs 10 and 12 by deleting the *Hin*dIII fragment spanning the first four codons of the *SUC2* ORF and a short portion of the polylinker sequence between the 3′-end of the $P_{ADH1}$ promoter and the *SUC2* ORF. As a result, the translation of $C_{ub}$ fusions 15 and 16 began at the first codon of the mature invertase, skipping its signal sequence. The $C_{ub}$ fusion 14 (Figure 2) was produced through an in-frame fusion of a PCR fragment containing the complete *TPI1* coding sequence and $C_{ub}$-Dha. The sequence between *TPI1* and $C_{ub}$ is as follows: AAC GGG TCG ACC GAC TAC AAG GAC GAC GAT GAC AAG GGC TCG ACC ATGTCG GGG GGG ATC CCT. The underlined sequences indicate, respectively, the last codon of *TPI1* and the first three codons of $C_{ub}$.

A fragment encoding $N_{ub}$-Sec62p was constructed using PCR amplification of a 1050 base pair (bp) fragment containing the *SEC62* ORF. PCR introduced a *Bam*HI site and a linker sequence in front of the start codon of *SEC62* and an *Xho*I site 173 bp downstream of the stop codon. The $P_{CUP1}$-$N_{ub}$ modules were cloned as *Bam*HI fragments in frame with the *SEC62* ORF. The sequence between $N_{ub}$ and *SEC62* is GGG ATC CCT TCT GGG ATG. The first three codons encode residues 35, 36, and 37 of $N_{ub}$, followed by the Gly-Ser linker and the start codon of *SEC62*. The *Bam*HI site is underlined. The final constructs resided in pRS316 or pRS313. $N_{ub}$-*TPI1*, $N_{ub}$-*SED5*, $N_{ub}$-*STE14* (a gift from N. Lewke), and $N_{ub}$-Sec62(ΔC60)-Dha were constructed similarly to $N_{ub}$-*SEC62*. With the exception of $N_{ub}$-Sec62(ΔC60)-Dha, which was placed in pRS316 and pRS313, all of these fusions resided in pRS314. The linker connecting codon 35 of $N_{ub}$ and the first codon of a linked gene was GGG ATC CCT GGG GAT ATG for $N_{ub}$-*TPI1* and $N_{ub}$-*SED5*, and GGG ATC CCT GGG GAT CAC for $N_{ub}$-*STE14*. Underlined are the *Bam*HI site and the first codon of the linked gene. The sequence GGG TCG ACC TTA ATG CAG AGA TCT GGC ATC ATG GTT connected the last codon

**Table 1.** Yeast strains

| Strain | Relevant genotype | Source/comment |
|---|---|---|
| YPH500 | *MATα ade2-101 his3-Δ200 leu2-Δ1 lys2-801 trp1-Δ63 ura3-52* | Sikorski and Hieter (1989) |
| JD53 | *MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52* | Dohmen *et al.* (1995) |
| RSY529 | *MATα his4 leu2-3,112 ura3-52 sec62-1* | Deshaies and Schekman (1989) |
| NJY51 | *MATα ade2-101 his3-Δ200 leu2-Δ1 lys2-801 trp1-Δ63 ura3-52 NUB-BOS1::pRS306* | Derivative of YPH500 |
| NJY62-1 | *MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUB-SEC62::pRS306* | Derivative of JD53 |
| NJY73-I | *MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUB-BOS1::pRS303* | Derivative of JD53 |
| NJY74-I | *MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUB-BOS1::pRS306* | Derivative of JD53 |
| NJY61-I | *MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUB-SEC61::pRS304* | Derivative of JD53 |
| NJY61-A | *MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUA-SEC61::pRS304* | Derivative of JD53 |
| NJY61-G | *MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUG-SEC61::pRS304* | Derivative of JD53 |

of *SEC62* in $N_{ub}$-Sec62(ΔC60)-Dha (codon 223, underlined) to the first two codons of DHFR (underlined).

$N_{ub}$-*BOS1* was constructed in part by PCR amplification, with two synthetic oligos and yeast genomic DNA as a template, yielding a 258-bp fragment containing the first 229 bp of the *BOS1* ORF. Upstream of the *BOS1* ATG was a short linker sequence and a *Bam*HI site, to allow in-frame fusion of the $P_{CUP1}$ promoter-$N_{ub}$ module. The sequence between $N_{ub}$ and *BOS1* reads: GGG ATC CCT CCA GGA ATG. The first four triplets encode residues 35, 36, 37, and 38 of $N_{ub}$, followed by the Gly codon and the start codon of *BOS1*. The *Bam*HI site is underlined. The 3′-region of the resulting fragment terminated in a *Sal*I site for insertion into the integrating vectors pRS306 or pRS303. The vector was cut at the unique *Eco*RI site in the *BOS1*-containing fragment and transformed into *S. cerevisiae* strains YPH500 and JD53 to produce, through homologous recombination, the integrated cassette that expressed $N_{ub}$-Bos1p from the $P_{CUP1}$ promoter. The presence of the desired gene fusion and the absence of wild-type *BOS1* were verified by PCR.

An integrated copy of $P_{CUP1}$-$N_{ub}$-*SEC62* was produced by amplifying the first 438 bp of the *SEC62* ORF, and then cloning it, using the *Bam*HI and *Eco*RI restriction sites, in frame behind the pRS306-$P_{CUP1}$-$N_{ub}$ cassette. A unique *Afl*II site in the *SEC62* ORF was used to linearize the plasmid for transformation and integration at the *S. cerevisiae SEC62* gene, yielding the strain NJY62-I. The N-terminal 147-residue fragment of Sec62p that was coexpressed with $N_{ub}$-Sec62p in the resulting strain has previously been shown to be inactive in translocation (Deshaies and Schekman, 1990). $N_{ub}$-*SEC61* was constructed by targeted integration of a $N_{ub}$-*SEC61*–containing fragment into *SEC61* of the *S. cerevisiae* strain JD53 (Table 1). Specifically, a fragment containing the first 875 bp of the *SEC61* ORF was amplified by PCR and inserted downstream of the pRS304- or pRS303-based $P_{CUP1}$-$N_{ub}$ cassette, using the flanking *Bam*HI and *Eco*RI sites. The linker sequence between $N_{ub}$ and *SEC61* was GGG ATC CCT GGG TCT GGG ATG. Underlined are the *Bam*HI site and the first codon of *SEC61*. For targeted integration, the plasmid was linearized at the unique *Stu*I site in the *SEC61* ORF to create the yeast NJY61-I. A detailed description of the NJY61 strains (Table 1) will be presented elsewhere (Wittke and Johnsson, unpublished data).

All of the $P_{CUP1}$ promoter-controlled ORFs were expressed under noninducing conditions (no copper added to the medium), except in the experiment shown in Figure 5B, where cells were incubated in the presence of 0.1 mM $CuSO_4$.

### *Immunoblotting*

Proteins fractionated by SDS-12.5% PAGE were electroblotted onto nitrocellulose (Schleicher & Schuell, Dassel, Germany) or polyvinylidene difluoride (Machery-Nagel, Düren, Germany) membranes, using the semidry transfer system (Hoeffer Pharmacia Biotech, San Francisco, CA). Blots were incubated with an anti-ha monoclonal antibody (Babco, Richmond, CA) and visualized using horseradish peroxidase-coupled goat anti-mouse antibody (Bio-Rad, Hercules, CA), the chemiluminescence detection system (Boehringer, Mannheim, Germany), and x-ray films. Where indicated, quantification was performed using the Lumi Imager system (Boehringer).

### *Pulse-Chase Analysis*

Yeast-rich (YPD) and synthetic minimal media with 2% dextrose (SD) were prepared as described previously (Dohmen *et al.*, 1995). *S. cerevisiae* cells expressing the $N_{ub}$ and $C_{ub}$ fusions were grown at 30°C in 10 ml of SD medium without externally added copper to an $OD_{600}$ of ~1 and labeled for 5 min with Redivue Promix-[$^{35}$S] (Amersham, Buckinghamshire, United Kingdom), followed (either directly or after a chase) by immunoprecipitation with the anti-ha monoclonal antibody, essentially as described by Johnsson and Varshavsky (1994a,b). The EndoH analysis of glycosylated proteins was carried out as described by Orlean *et al.* (1991). Samples were concentrated before SDS-12.5% PAGE by precipitation with chloroform/methanol. Gels were fixed and enhanced for fluorography. For quantitative analysis, a dried gel was exposed and scanned using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA).

## RESULTS

### *Experimental Strategy*

The use of split Ub to monitor the proximity between the proteins X and Y requires the construction of two "complementary" fusion proteins. One fusion bears $N_{ub}$ (see INTRODUCTION) linked to X ($N_{ub}$-X) and the other bears $C_{ub}$ linked to both Y and a reporter protein R at the C terminus of $C_{ub}$ (Y-$C_{ub}$-R). The liberation of the reporter through the Ub-dependent cleavage by UBPs indicates the in vivo reconstitution

of a quasi-native Ub from $N_{ub}$ and $C_{ub}$. In the split-Ub assay, the efficiency of cleavage at the C terminus of $C_{ub}$ in Y-$C_{ub}$-R is measured relative to the efficiency of cleavage observed with selected reference (control) proteins (Figure 1).

To monitor protein interactions during translocation of a protein across the ER membrane, $N_{ub}$ was fused to the N terminus of a membrane protein that is a part of the translocation machinery (Figure 1). Owing to the constraint of the assay, which requires the cytosolic location of the reconstituted Ub, the N terminus of this membrane protein must be located in the cell's cytosol. Sec62p has an N-terminal cytosolic domain of 158 residues, which is followed by two membrane-spanning segments and a C-terminal segment also facing the cytosol (Deshaies and Schekman, 1990). $N_{ub}$ was therefore fused to the N terminus of Sec62p, yielding $N_{ub}$-Sec62p. $C_{ub}$ was sandwiched between the 56 N-terminal residues of the precursor of *S. cerevisiae* $\alpha$-factor pheromone (prepro-$\alpha$-factor) and the ha epitope-tagged mouse dihydrofolate reductase (DHFR-ha; denoted as Dha) as a reporter protein, yielding Mf$\alpha_{37}$-$C_{ub}$-Dha (Figure 2). The cleavage of the $C_{ub}$-containing fusion at the $C_{ub}$–Dha junction was detected with a monoclonal anti-ha antibody.

### *Split-Ub Detects a Proximity between a Translocating Protein and Sec62p*

We first verified that Mf$\alpha_{37}$-$C_{ub}$-Dha could be translocated across the ER membrane and that the N-terminal extension of Sec62p with $N_{ub}$ did not interfere with the Sec62p function in translocation. After a 5-min pulse of wild-type *S. cerevisiae* with $^{35}$S-methionine, the labeled Mf$\alpha_{37}$-$C_{ub}$-Dha was immunoprecipitated as a glycosylated and unclipped fusion (Figure 3A). Thus, Mf$\alpha_{37}$-$C_{ub}$-Dha could indeed be translocated into the lumen of ER. Introduction of the same Mf$\alpha_{37}$-$C_{ub}$-Dha construct into the yeast strain RSY529, which carries a temperature-sensitive (ts) variant of Sec62p (Rothblatt *et al.*, 1989), confirmed the severe translocation defect of this strain. About 50% of the pulse-labeled Mf$\alpha_{37}$-$C_{ub}$-Dha entered the lumen of the ER in this strain at the semipermissive temperature of 30°C, while the rest remained in the cytosol (Figure 3A). Thus, the translocation of Mf$\alpha_{37}$-$C_{ub}$-Dha depends on Sec62p. This made it possible to determine whether $N_{ub}$-Sec62p is functionally active. The test utilized $N_{ug}$-Sec62p, in which the N-terminal half of Ub contained Gly-13 instead of wild-type Ile-13. This derivative, denoted as $N_{ug}$, has a lower affinity for $C_{ub}$ than the wild-type $N_{ub}$ (Johnsson and Varshavsky, 1994a). We chose $N_{ug}$-Sec62p for this experiment to minimize the reconstitution of the Ub moiety through interactions between $N_{ub}$-Sec62p and potentially arrested molecules of Mf$\alpha_{37}$-$C_{ub}$-Dha, which might be localized in the cytosol. Plasmids expressing $N_{ug}$-Sec62p and

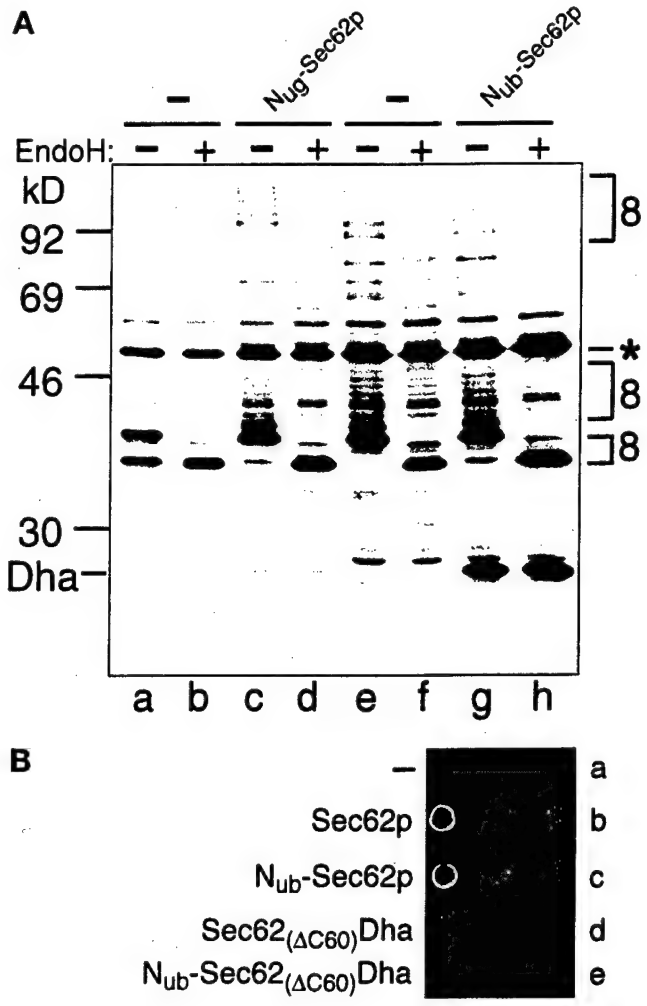



**Figure 3.** Sec62p is close to the signal sequence of the $\alpha$-factor precursor. (A) *S. cerevisiae* cells expressing Mf$\alpha_{37}$-$C_{ub}$-Dha (construct 8; Figure 2) were labeled for 5 min with $^{35}$S-methionine. The extracted proteins were immunoprecipitated with anti-ha antibody, followed by a mock treatment (lanes a, c, e, and g) or the treatment with EndoH (lanes b, d, f, and h), and SDS-PAGE. The results with cells coexpressing $N_{ug}$- or $N_{ub}$-Sec62p are shown in lanes c and d and g and h, respectively. The analysis was performed with $N_{ug}$-Sec62p in the *S. cerevisiae* mutant RSY529 carrying a ts allele of *SEC62* (lanes a–d) or with $N_{ub}$-Sec62p in the wild-type yeast (lanes e–h). Number 8 (following the numbering of the constructs in Figure 2) on the right indicates the positions of uncleaved Mf$\alpha_{37}$-$C_{ub}$-Dha and its glycosylated forms. An asterisk denotes an unrelated yeast protein that cross-reacts with the anti-ha antibody. (B) $N_{ub}$-Sec62p encodes a functionally active protein. RSY529 cells carrying an empty plasmid (a), Sec62p (b), $N_{ub}$-Sec62p (c), Sec62(ΔC60)Dha (d), or $N_{ub}$-Sec62(ΔC60)Dha (e) were spotted on minimal media and grown for 2 d at 30°C (semipermissive temperature for unmodified RSY529).

Mf$\alpha_{37}$-$C_{ub}$-Dha were cotransformed into RSY529 cells and assayed at 30°C. As in wild-type cells, only translocated Mf$\alpha_{37}$-$C_{ub}$-Dha, but virtually no free Dha or nontranslocated Mf$\alpha_{37}$-$C_{ub}$-Dha, was detected after immunoprecipitation and EndoH treatment of the

cells that had been labeled for 5 min with $^{35}$S-methionine (Figure 3A).

To test $N_{ub}$-Sec62p directly ($N_{ub}$ is the wild-type half of Ub, containing Ile at position 13), we examined its ability to complement the growth defect of RSY529 cells. RSY529 cells expressing $N_{ub}$-Sec62p were found to grow at the semipermissive temperature of 30°C, in contrast to congenic cells carrying a control plasmid (Figure 3B). To verify that the suppression of the ts phenotype was not due to the initiation of translation from the first (internal) ATG codon of Sec62p within the $N_{ub}$-Sec62p fusion, the rescue experiment was successfully repeated with the otherwise identical derivative of $N_{ub}$-Sec62p that lacked the first ATG of *SEC62* (our unpublished results).

A significant amount of free Dha was generated when $Mf\alpha_{37}$-$C_{ub}$-Dha was expressed (in either wild-type or RSY529 cells) together with $N_{ub}$-Sec62p, which contained the wild-type half of Ub (Figure 3A and our unpublished results). We concluded that Sec62p is close to the nascent polypeptide chain during its translocation into the ER. The cleavage at the C terminus of $C_{ub}$ requires its interaction with $N_{ub}$ and depends on the presence of UBPs (Johnsson and Varshavsky, 1994a). Since UBPs have previously been shown to be absent from the ER (Johnsson and Varshavsky, 1994b), the free Dha moiety had to be produced in the cytosol. Fractionation experiments confirmed that free Dha was absent from membrane-enclosed compartments in whole-cell extracts (our unpublished results). An entirely independent evidence for this conclusion was produced by replacing Dha in $Mf\alpha_{37}$-$C_{ub}$-Dha with Ura3p as the reporter moiety. Ura3p confers the Ura$^+$ phenotype on *ura3*Δ cells only if Ura3p has access to the cytosol (Johnsson and Varshavsky, 1994b). In our tests, the cytosolic Ura3p was produced only if $Mf\alpha_{37}$-$C_{ub}$-Ura3p was coexpressed with $N_{ub}$-Sec62p (compare A and B in Figure 7), in agreement with the other evidence (see above) that the cleavage at the $C_{ub}$–protein junction takes place exclusively in the cytosol.

The transient nature of the proximity between Sec62p and the nascent chain of a translocated protein was indicated by the near-absence of the released Dha moiety if $Mf\alpha_{37}$-$C_{ub}$-Dha was coexpressed with either $N_{ug}$-Sec62p or $N_{ua}$-Sec62p instead of $N_{ub}$-Sec62p ($N_{ua}$ denotes Ala at position 13 of $N_{ub}$); by contrast, the same experiment with $N_{ub}$-Sec62p resulted in a significant cleavage of $Mf\alpha_{37}$-$C_{ub}$-Dha (Figures 3A and 6C). Previous work (Johnsson and Varshavsky, 1994a) has shown that $N_{ua}$ and $N_{ug}$ can induce significant Ub reconstitution when either of them and $C_{ub}$ are linked to polypeptides that form a stable (long-lived) complex in a cell. In summary, the observed absence of significant Ub reconstitution with $N_{ua}$ and $N_{ug}$ (in contrast to $N_{ub}$) was interpreted to signify a close but transient (short-lived) proximity between Sec62p and $Mf\alpha_{37}$-$C_{ub}$-Dha.

### *Specificity of the Spatial Proximity between a Signal Sequence-bearing Nascent Polypeptide and Sec62p*

A commonly used negative control in a translocation assay is a protein with a defective or absent signal sequence (Allison and Young, 1988; Müsch *et al.*, 1992). Such a control is not entirely compatible with spatio-temporal aspects of the split-Ub assay. Specifically, a $C_{ub}$-fusion protein lacking a signal sequence accumulates in the cytosol (where the split-Ub assay operates), whereas an analogous signal sequence-bearing protein is continuously removed from this compartment. A direct comparison between reactions that involve a signal sequence-bearing polypeptide and its signal sequence-lacking counterpart requires the ability to compare the local concentrations of the two polypeptides at the site of translocation. We are not aware of an in vivo technique that would be independent of the split-Ub assay and at the same time would allow a measurement of these parameters. Therefore, we devised an alternative control. The extent of cleavage of $Mf\alpha_{37}$-$C_{ub}$-Dha at the $C_{ub}$–Dha junction should reflect the time-averaged spatial proximity between the nascent Mfα-chain and a coexpressed $N_{ub}$-containing fusion. By comparing the extent of cleavage of $Mf\alpha_{37}$-$C_{ub}$-Dha in the presence of $N_{ub}$-Sec62p (Figure 3A) with the analogous activity of $N_{ub}$-fusion proteins that are not involved in the ER targeting and translocation, we could assess the specificity of the reaction between $N_{ub}$-Sec62p and $Mf\alpha_{37}$-$C_{ub}$-Dha.

Four $N_{ub}$-fusion proteins, $N_{ub}$-Bos1p, $N_{ub}$-Ste14p, $N_{ub}$-Sed5p, and $N_{ub}$-Tpi1p were tested in the split-Ub assay with $Mf\alpha_{37}$-$C_{ub}$-Dha. The expected intracellular locations of these $N_{ub}$ fusions, and their predicted topologies in the membrane are shown in Figure 2A (Shim *et al.*, 1991; Banfield *et al.*, 1994; Sapperstein *et al.*, 1994; Lewke and Johnsson, unpublished data). We found that, in contrast to $N_{ub}$-Sec62p, none of the four tested $N_{ub}$ fusions induced a significant cleavage of $Mf\alpha_{37}$-$C_{ub}$-Dha (Figure 4A). A small amount of free Dha could be detected in the immunoblots when $N_{ub}$-Bos1p was overexpressed. The lack of significant Ub reconstitution from $N_{ub}$ and $C_{ub}$ upon coexpression of $Mf\alpha_{37}$-$C_{ub}$-Dha and the $N_{ub}$-modified ER membrane proteins, Bos1p, Ste14p (Figure 4A), and Sec12p (Nakano *et al.*, 1988; our unpublished results), confirmed that the steady-state concentration of $Mf\alpha_{37}$-$C_{ub}$-Dha in the cytosol was extremely low.

To verify that the observed absence of Ub reconstitution (Figure 4A) was not due to either low concentrations of the tested fusion proteins or reduced accessibility of their linked $N_{ub}$ moieties, we compared the activity of these $N_{ub}$ fusions toward a cytosolic $C_{ub}$-fusion protein. $C_{ub}$-Dha was fused to the C terminus of the cytosolic enzyme triosephosphate isomerase

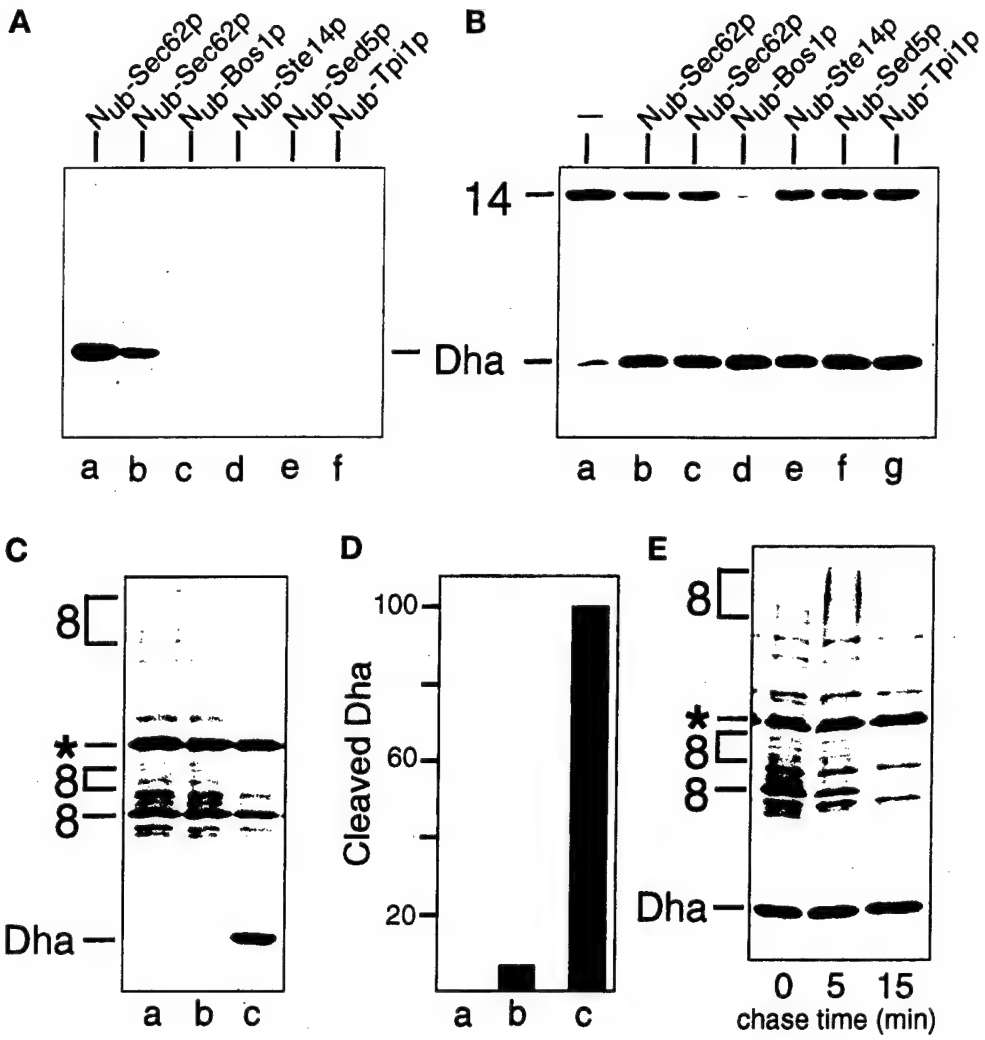



**Figure 4.** The in vivo proximity between Sec62p and Mf$\alpha_{37}$-$C_{ub}$-Dha is transient and specific. (A) Immunoblot analysis of extracts of *S. cerevisiae* coexpressing the Mf$\alpha_{37}$-$C_{ub}$-Dha (construct 8; Figure 2) and one of the following constructs: $N_{ub}$-Sec62p, integrated (lane a) or plasmid-borne (lane b); $N_{ub}$-Bos1p (lane c); $N_{ub}$-Ste14p (lane d); $N_{ub}$-Sed5p(lane e); and $N_{ub}$-Tpi1p (lane f). (B) Immunoblot analysis of extracts of *S. cerevisiae* expressing the Tpi1p-$C_{ub}$-Dha fusion (construct 14; Figure 2) alone (lane a) or together with one of the following constructs: $N_{ub}$-Sec62p, either integrated (lane b) or plasmid-borne (lane c); $N_{ub}$-Bos1p (lane d); $N_{ub}$-Ste14p (lane e); $N_{ub}$-Sed5p(lane f); and $N_{ub}$-Tpi1p (lane g). Number 14 on the left indicates the position of uncleaved Tpi1p-$C_{ub}$-Dha. (C) *S. cerevisiae* cells expressing Mf$\alpha_{37}$-$C_{ub}$-Dha (construct 8; Figure 2) together with either the vector (lane a), $N_{ub}$-Bos1p (lane b), or $N_{ub}$-Sec62p (lane c) were labeled for 5 min with $^{35}$S-methionine. The extracted proteins were immunoprecipitated with anti-ha antibody and analyzed by SDS-PAGE. (D) Quantitation of the pulse-labeling experiment (C) using PhosphorImager. The extent of Dha release in the presence of $N_{ub}$-Sec62p was arbitrarily set at 100. The averages of three experiments are shown. Lanes a, b, and c are the same as in panel C. (E) *S. cerevisiae* cells expressing Mf$\alpha_{37}$-$C_{ub}$-Dha together with $N_{ub}$-Sec62p were labeled for 5 min with $^{35}$S-methionine and chased for 5 and 15 min, followed by extraction of proteins, immunoprecipitation with anti-ha antibody, and SDS-PAGE.

(Tpi1p), yielding Tpi1-$C_{ub}$-Dha (Figure 2B). All of the $N_{ub}$-fusion proteins in Figure 2A induced a significant release of Dha from the test protein Tpi1-$C_{ub}$-Dha (Figure 4B). This analysis also suggested that $N_{ub}$-Bos1p was expressed to higher levels than other $N_{ub}$ fusions.

To quantify the relative proximities of $N_{ub}$-Bos1p and $N_{ub}$-Sec62p to Mf$\alpha_{37}$-$C_{ub}$-Dha, yeast cells were labeled for 5 min with $^{35}$S-methionine, and the released Dha was determined as described in the legend to Figure 4. Coexpression of $N_{ub}$-Sec62p and Mf$\alpha_{37}$-$C_{ub}$-Dha yielded ~15 times more of the free Dha than coexpression of $N_{ub}$-Bos1p and Mf$\alpha_{37}$-$C_{ub}$-Dha (Figure 4, C and D). Assuming that the $N_{ub}$ moieties in $N_{ub}$-Sec62p and $N_{ub}$-Bos1p were equally accessible to the cytosol (Figure 4B), we concluded that the time-averaged proximity between the nascent chain of Mf$\alpha_{37}$-$C_{ub}$-Dha and the $N_{ub}$-bearing transmembrane proteins was much higher for Sec62p than for the ER membrane proteins that are not involved in targeting or translocation. Note that this analysis may actually underestimate the proximity of Sec62p to the nascent chain, because we invariably observed a more efficient cleavage of Mf$\alpha_{37}$-$C_{ub}$-Dha when $N_{ub}$-Sec62p was the only form of Sec62p in the cell (Figures 4A and 6A). Therefore we interpret the reduced cleavage of Mf$\alpha_{37}$-$C_{ub}$-Dha in the presence of both $N_{ub}$-Sec62p and the native Sec62p as the consequence of competition between those two Sec62p-containing species for either the signal sequences of translocated proteins or the ligands of Sec62p in the complex of Sec proteins.

Recent evidence indicates that misfolded or otherwise abnormal proteins in the lumen of the ER can be retrotransported across the ER membrane back into

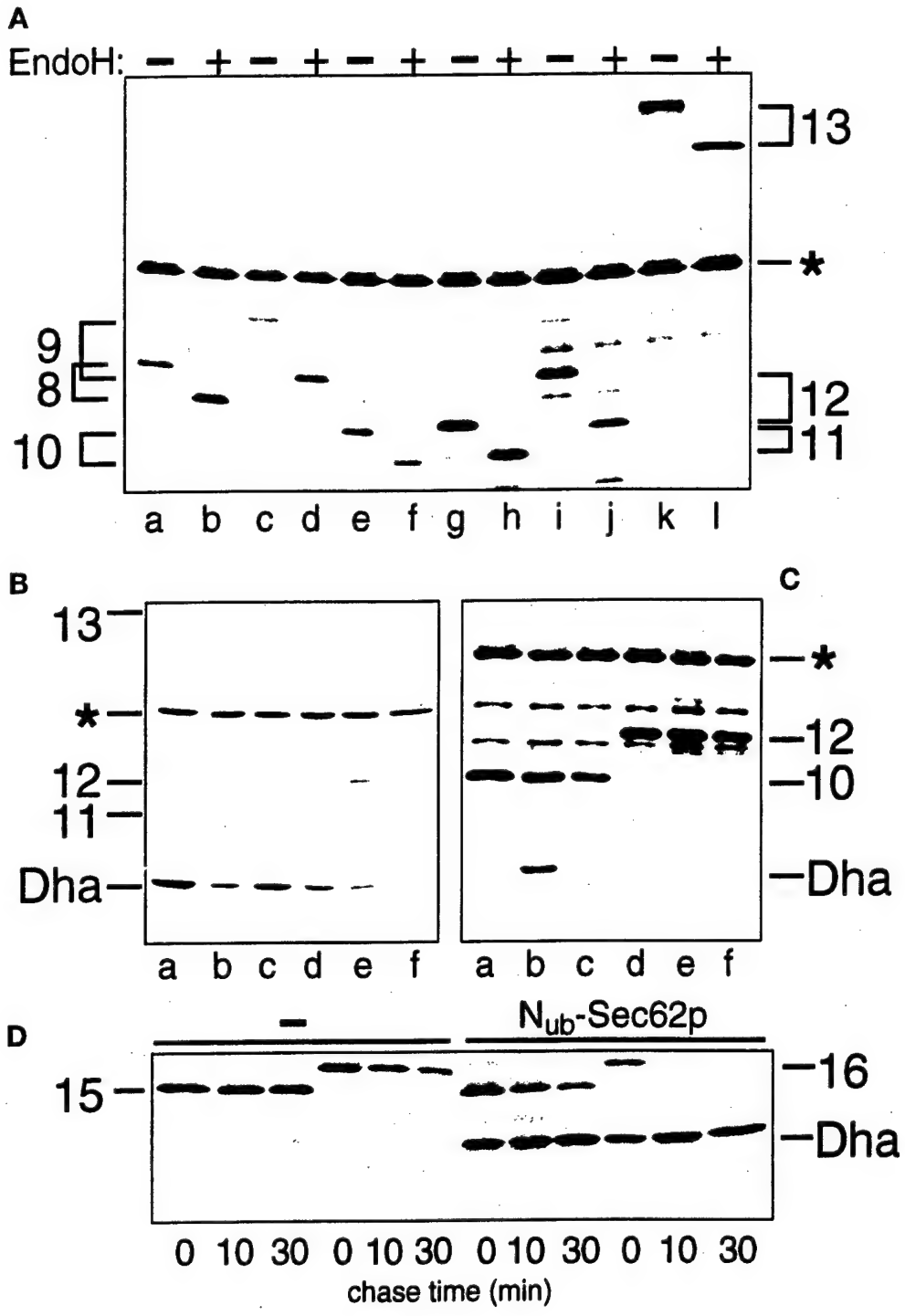



**Figure 5.** The nature of the signal sequence and its distance from $C_{ub}$ determine the extent of cleavage of $C_{ub}$-R in the presence of $N_{ub}$-Sec62p. (A) *S. cerevisiae* expressing $Mf\alpha_{37}$-$C_{ub}$-Dha (construct 8; Figure 2) (lanes a and b), $Mf\alpha_{65}$-$C_{ub}$-Dha (construct 9) (lanes c and d), $Suc2_{23}$-$C_{ub}$-Dha (construct 10) (lanes e and f), $Suc2_{33}$-$C_{ub}$-Dha (construct 11) (lanes g and h), $Suc2_{59}$-$C_{ub}$-Dha (construct 12) (lanes i and j) and $Suc2_{518}$-$C_{ub}$-Dha (construct 13) (lanes k and l) were labeled with $^{35}$S-methionine for 5 min. The extracted proteins were either mock-treated (lanes a, c, e, g, i, and k) or treated with EndoH (lanes b, d, f, h, j, and l), followed by immunoprecipitation with anti-ha antibody and SDS-PAGE. (B) Same as panel A but the cells also contained $N_{ub}$-Sec62p in addition to the Cub-fusions $Mf\alpha_{37}$-$C_{ub}$-Dha, $Mf\alpha_{65}$-$C_{ub}$-Dha, $Suc2_{23}$-$C_{ub}$-Dha, $Suc2_{33}$-$C_{ub}$-Dha, $Suc2_{59}$-$C_{ub}$-Dha, $Suc2_{518}$-$C_{ub}$-Dha (lanes a–f). The analysis was carried out by immunoblotting whole- cell extracts with the anti-ha antibody. (C) *S. cerevisiae* cells expressing $Suc2_{23}$-$C_{ub}$-Dha (construct 10; Figure 2) (lanes a–c) and $Suc2_{59}$-$C_{ub}$-Dha (construct 12; Figure 2) (lanes d–f) together with either $N_{ub}$-Sec62p (lanes b and e), $N_{ub}$-Bos1p (lanes c and f) or the vector (lanes a and d) were labeled for 5 min with $^{35}$S-methionine. Whole-cell extracts were immunoprecipitated with anti-ha antibody, followed by SDS-PAGE and autoradiography. (D) *S. cerevisiae* cells expressing $\Delta Suc2_{23}$-$C_{ub}$-Dha (construct 15; Figure 2) or $\Delta Suc2_{59}$-$C_{ub}$-Dha (construct 16; Figure 2) together with either the vector (first six lanes) or $N_{ub}$-Sec62p (last six lanes) were labeled for 5 min with $^{35}$S-methionine and chased for 10 and 30 min, followed by extraction, immunoprecipitation with anti-ha antibody, and SDS-PAGE. Numbers 15 and 16 indicate the positions of the corresponding (uncleaved) $C_{ub}$ fusions.

the cytosol, where they are degraded by the Ub system (Biederer *et al.*, 1996; Hiller *et al.*, 1996; Wiertz *et al.*, 1996). This retrotransport involves at least some of the components of the known ER translocation machinery (Plemper *et al.*, 1997). To determine whether the cleavage of $Mf\alpha_{37}$-$C_{ub}$-Dha at the $C_{ub}$-Dha junction occurs during translocation into the ER or during (in this case) a hypothetical retrotransport from the ER, cells coexpressing $N_{ub}$-Sec62p and $Mf\alpha_{37}$-$C_{ub}$-Dha were labeled for 5 min with $^{35}$S-methionine, and then chased for 15 min (Figure 4E). Although the translocated $Mf\alpha_{37}$-$C_{ub}$-Dha disappeared rapidly during the chase,

the amount of free Dha that accumulated during the pulse remained constant.

We conclude that the in vivo proximity between Sec62p and Mf$\alpha_{37}$-$C_{ub}$-Dha that is detected by the split-Ub assay occurs either during or very shortly after the synthesis of Mf$\alpha_{37}$-$C_{ub}$-Dha. The apparent disappearance of the pulse-labeled, translocated Mf$\alpha_{37}$-$C_{ub}$-Dha during the chase accounts for the difficulty in detecting this species by a steady-state assay such as immunoblotting (Figures 4A, 5B, and 6A). The likely cause of the disappearance of translocated Mf$\alpha_{37}$-$C_{ub}$-Dha is its molecular mass heterogeneity, owing to its glycosylation, which results in a smear upon SDS-PAGE (Figures 3A, 4C, and 4E).

### *The Efficiency of Ub Reconstitution Mediated by $N_{ub}$-Sec62p Depends on Both the Identity of a Signal Sequence and the Position of $C_{ub}$ in the Nascent Polypeptide Chain*

The proximity of Sec62p to the signal sequence of Mf$\alpha_{37}$-$C_{ub}$-Dha is detected, in the split-Ub assay, through the ability of $N_{ub}$-Sec62p to induce the cleavage of Mf$\alpha_{37}$-$C_{ub}$-Dha at the $C_{ub}$–Dha junction (Figure 3A). If this cleavage reflects the physical proximity between Sec62p and a signal sequence, the efficiency of cleavage should decrease if the $C_{ub}$ moiety is moved closer to the C terminus of the nascent polypeptide chain. However, this purely spatial consideration neglects the temporal aspect of the translocation process (Walter and Johnson, 1994). The targeting and the actual translocation are initiated during or shortly after the synthesis of a signal sequence-bearing protein. Consequently, the C-terminal parts of the nascent chain may still be synthesized, or at least associated with the ribosome, at the time when Sec62p and the signal sequence have already become spatially close. Extending the spacer would increase the distance between $C_{ub}$ and the signal sequence of Mf$\alpha_{37}$-$C_{ub}$-Dha. This would be expected to decrease the time window available for the interaction between the $C_{ub}$ moiety of Mf$\alpha_{37}$-$C_{ub}$-Dha and the $N_{ub}$ moiety of $N_{ub}$-Sec62p. Therefore, a test of this kind cannot deconvolute the contribution of each of the two parameters (increased spatial distance along the chain between Sec62p and $C_{ub}$ and decreased time window for the $N_{ub}$–$C_{ub}$ interaction) to the overall effect of extending the length of the polypeptide between the signal sequence and $C_{ub}$. These constraints notwithstanding, moving the $C_{ub}$ moiety of Mf$\alpha_{37}$-$C_{ub}$-Dha further away from its signal sequence makes it possible to gauge the accessibility of Sec62p to specific regions of the nascent polypeptide chain in vivo.

In the actual experiment, the distance between the signal sequence of Mf$\alpha_{37}$-$C_{ub}$-Dha and its $C_{ub}$ moiety was increased from 37 to 65 residues (Mf$\alpha_{65}$-$C_{ub}$-Dha; Figure 2B, construct 9). The results of EndoH treatment of Mf$\alpha_{65}$-$C_{ub}$-Dha immunoprecipitated from pulse-labeled wild-type cells confirmed that Mf$\alpha_{65}$-$C_{ub}$-Dha was efficiently translocated into the ER (Figure 5A). However, the efficiency of the Dha-yielding cleavage of Mf$\alpha_{65}$-$C_{ub}$-Dha upon coexpression of $N_{ub}$-Sec62p was clearly reduced in comparison to the same cleavage with Mf$\alpha_{37}$-$C_{ub}$-Dha and $N_{ub}$-Sec62p (Figure 5B).

Both the kinetics and the mode of targeting for translocation are influenced by the identity of a signal sequence (Bird *et al.*, 1987; Johnsson and Varshavsky, 1994b; Ng *et al.*, 1996). For example, the efficient translocation of invertase (Suc2p) requires the SRP, in contrast to a much weaker requirement for SRP in the case of the prepro-$\alpha$-factor's signal sequence (Hann and Walter, 1991; Ogg *et al.*, 1992; Johnsson and Varshavsky, 1994b). Consequently, the coupling between translation and translocation is tighter for proteins bearing the invertase signal sequence than for proteins carrying the signal sequence of the $\alpha$-factor.

We assessed the in vivo proximity of the invertase signal sequence to Sec62p by measuring the reconstitution of Ub from $N_{ub}$-Sec62p and Suc2-$C_{ub}$-Dha, where the Suc2p moiety was linked to $C_{ub}$ through a spacer of increasing length (Figure 2B). The expression and efficient translocation of different Suc2-$C_{ub}$-Dha constructs were assayed by immunoprecipitation and subsequent EndoH treatment (Figure 5A). The proximity of $C_{ub}$ in Suc2-$C_{ub}$-Dha to $N_{ub}$ of $N_{ub}$-Sec62p was assayed by immunoblot detection of the cleavage-derived free Dha in whole-cell extracts. The pattern already observed for the Mf$\alpha$-$C_{ub}$-Dha constructs recurred with the constructs bearing the invertase signal sequence (Figure 5B). Moreover, coexpression of $N_{ub}$-Sec62p with either $Suc2_{23}$-$C_{ub}$-Dha or $Suc2_{33}$-$C_{ub}$-Dha yielded lower amounts of free Dha than the analogous assays with $N_{ub}$-Sec62p and Mf$\alpha_{37}$-$C_{ub}$-Dha, which bears a spacer of comparable length (Figure 5B).

Pulse-chase analyses with cells expressing $N_{ub}$-Sec62p (or $N_{ub}$-Bos1p) and either $Suc2_{23}$-$C_{ub}$-Dha or $Suc2_{59}$-$C_{ub}$-Dha confirmed the immunoblot data. Specifically, a significant release of free Dha was observed only for the pair of $N_{ub}$-Sec62p and $Suc2_{23}$-$C_{ub}$-Dha (Figure 5C). Our previous work has shown that the segment of the nascent polypeptide chain where the $C_{ub}$ moiety was inserted in either the $Suc2_{23}$-$C_{ub}$-Dha or the Mf$\alpha_{37}$-$C_{ub}$-Dha fusion is transiently exposed to the cytosol—until the initiation of ER translocation (Johnsson and Varshavsky, 1994b). Therefore, we compared the ratios of cleaved to uncleaved $Suc2_{23}$-$C_{ub}$-Dha and Mf$\alpha_{37}$-$C_{ub}$-Dha. Cells expressing $N_{ub}$-Sec62p and either the $C_{ub}$ fusion 8 or 10 (Figure 2B) were labeled for 5 min with $^{35}$S-methionine and processed for immunoprecipitation with anti-ha antibody, followed by determination of the cleaved-to-uncleaved ratio (Figure 6D). This ratio, a measure of the time-averaged proximity of Sec62p to a translocating protein,

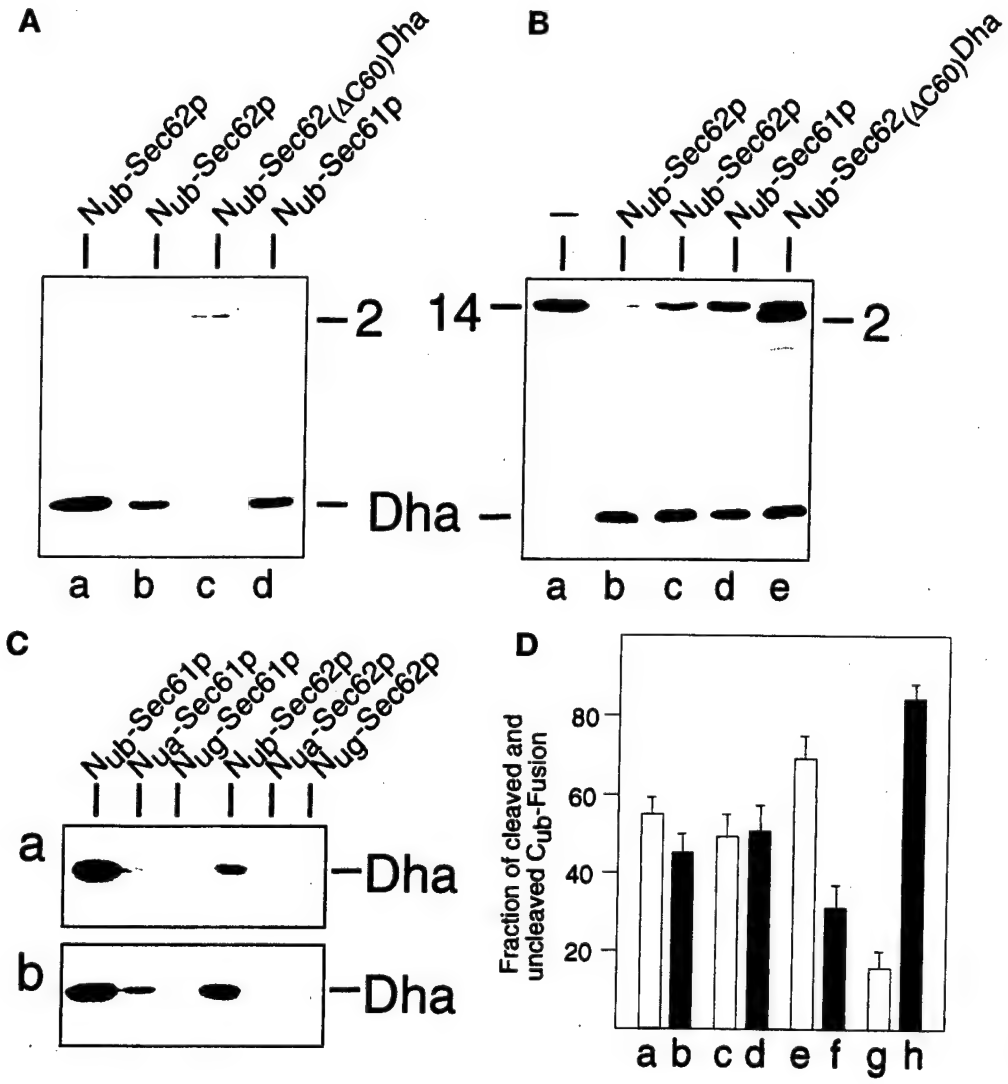



**Figure 6.** Sec61p, but not a mutant of Sec62p, are close to the nascent chain of a translocated protein. (A) These assays employed *S. cerevisiae* expressing $Mf\alpha_{37}$-$C_{ub}$-Dha (construct 8, Figure 2) and one of the following $N_{ub}$ fusions (Figure 2): $N_{ub}$-Sec62p, either integrated (lane a) or plasmid borne (lane b); $N_{ub}$-Sec62(ΔC60)Dha (lane c); and $N_{ub}$-Sec61p (lane d). Whole-cell extracts from these strains were subjected to immunoblot analysis with anti-ha antibody. (B) Same as panel A but the same $N_{ub}$ fusions were coexpressed with Tpi1-$C_{ub}$-Dha (construct 14; Figure 2). Numbers 2 and 14 indicate the positions of the corresponding (uncleaved) fusions. (C) Lane a: *S. cerevisiae* expressing $Suc2_{23}$-$C_{ub}$-Dha (construct 10; Figure 2) together with either $N_{ub}$-Sec61p, $N_{ua}$-Sec61p, $N_{ug}$-Sec61p, $N_{ub}$-Sec62p, $N_{ua}$-Sec62p, or $N_{ug}$-Sec62p; lane b: same as lane a but cells expressed $Mf\alpha_{37}$-$C_{ub}$-Dha (construct 8; Figure 2) instead of $Suc2_{23}$-$C_{ub}$-Dha. (D) *S. cerevisiae* cells expressing $Mf\alpha_{37}$-$C_{ub}$-Dha together with $N_{ub}$-Sec61p (a and b) or $N_{ub}$-Sec62p (e and f), and cells expressing $Suc2_{23}$-$C_{ub}$-Dha together with $N_{ub}$-Sec61p (c and d) or $N_{ub}$-Sec62p (g and h) were labeled for 5 min with $^{35}$S-methionine. Whole-cell extracts were immunoprecipitated with anti-ha antibody, followed by SDS-PAGE, and quantitation of the cleaved and uncleaved $C_{ub}$ fusions using PhosphorImager. Shown are the relative amounts of the cleaved (white bars: a, c, e, and g) and uncleaved (black bars: b, d, f, and h) $C_{ub}$ fusions. The sum of a cleaved and uncleaved fusion was set at 100 in each of the three independent experiments. SDs are also indicated.

was ~eightfold higher for a nascent polypeptide bearing the signal sequence of α-factor than for a nascent polypeptide bearing the invertase signal sequence (Figure 6D; compare Figures 4C and 5C).

Spacer sequences of different length or composition upstream of the $C_{ub}$ moiety might nonspecifically influence the interaction between $N_{ub}$ and $C_{ub}$. To assess this potential spacer effect, we constructed signal sequence-lacking versions of $Suc2_{23}$-$C_{ub}$-Dha and $Suc2_{59}$-$C_{ub}$-Dha (Figure 2B, $C_{ub}$ fusions 15 and 16), and compared their ability to reconstitute Ub in the presence of coexpressed $N_{ub}$-Sec62p. Both of these $C_{ub}$ fusions were cleaved at the $C_{ub}$-Dha junction at approximately the same rate in the presence of $N_{ub}$-Sec62p (Figure 5D), in contrast to the marked difference in the rate of cleavage observed for their signal sequence-bearing counterparts (Figure 5, B and C). This control experiment further emphasized the effect of distance between a signal sequence and the $C_{ub}$ moiety on the efficiency of Ub reconstitution in the presence of $N_{ub}$-Sec62p. We conclude that the accessibility of Sec62p in vivo to a specific region of the nascent polypeptide chain is influenced by both the nature of a signal sequence and its distance from that region.

### *Sec61p Is Equidistant from Two Different Signal Sequences*

A direct comparison between two different signal sequences upstream of the $C_{ub}$ moiety presumes approximately equal residence times of the corresponding $C_{ub}$ moieties in the cytosol. It is also essential to know that the influence of the identity of a signal sequence on the rate of Ub reconstitution is not due to

a nonspecific intramolecular interaction. One way to address these issues involves measuring the reconstitution of Ub from the $C_{ub}$ moieties of the fusions 8 and 10 (Figure 2B) and a $N_{ub}$-containing fusion that is not involved in translocation. As illustrated in Figure 4, this test is not feasible because of the rapid translocation of $C_{ub}$-containing constructs into the ER. Note, however, that since Sec61p is the central component of the translocation pore, proteins that utilize different targeting pathways will converge at Sec61p shortly before their translocation (Jungnickel and Rapoport, 1995). Taking advantage of this property of Sec61p, we assayed the proximity of $Mf\alpha_{37}$-$C_{ub}$-Dha and $Suc2_{23}$-$C_{ub}$-Dha to $N_{ub}$-Sec61p. If $Mf\alpha_{37}$-$C_{ub}$-Dha and $Suc2_{23}$-$C_{ub}$-Dha are cleaved at the $C_{ub}$–Dha junction equally well in the presence of $N_{ub}$-Sec61p, the above interpretation of the observed selectivity of $N_{ub}$-Sec62p toward $Mf\alpha_{37}$-$C_{ub}$-Dha (Figures 5 and 6) would be confirmed.

To carry out this test, $N_{ub}$ was fused to the cytosolic N terminus of Sec61p (Figure 2A) (Wilkinson *et al.*, 1996). $N_{ub}$-Sec61p is functionally active (Wittke and Johnsson, unpublished data). It induced the release of free Dha from of $Mf\alpha_{37}$-$C_{ub}$-Dha and Tpi1p-Cub-Dha with efficiency similar to that of $N_{ub}$-Sec62p (Figure 6, A and B). Thus, the split-Ub assay independently confirmed that Sec61p is close to the nascent polypeptide chain during its translocation. To compare the in vivo interactions of Sec61p with the $C_{ub}$ fusions 8 and 10, which bore different signal sequences (Figure 2B), the amount of free Dha was determined by immunoblotting of whole-cell extracts. It was found that in the presence of $N_{ub}$-Sec61p, similar amounts of Dha were released from $Mf\alpha_{37}$-$C_{ub}$-Dha and $Suc2_{23}$-$C_{ub}$-Dha, whereas in the presence of $N_{ub}$-Sec62p twice as much Dha was released from $Mf\alpha_{37}$-$C_{ub}$-Dha than from $Suc2_{23}$-$C_{ub}$-Dha (Figure 6C).

This result was confirmed and extended by labeling the cotransformed cells for 5 min with $^{35}$S-methionine and quantifying the ratio of cleaved-to-uncleaved $C_{ub}$ fusions (Figure 6D). As was already observed by the immunoblot analysis, the above ratio was ~1 for both $Mf\alpha_{37}$-$C_{ub}$-Dha and $Suc2_{23}$-$C_{ub}$-Dha in the presence of $N_{ub}$-Sec61p, but ~2 for $Mf\alpha_{37}$-$C_{ub}$-Dha, and ~0.25 for $Suc2_{23}$-$C_{ub}$-Dha in the presence of $N_{ub}$-Sec62p (Figure 6D). The difference revealed by the pulse-immunoprecipitation analysis is higher than the estimate obtained by the immunoblot analysis, most likely because of the continuous accumulation of cleaved (and long-lived) Dha before the processing of cells for immunoblotting.

### *A C-terminally Truncated Sec62p Is No Longer Proximal to the Signal Sequence*

Does the proximity of Sec62p to a nascent polypeptide chain that is detected by the split-Ub assay reflect the physical binding of the signal sequence to this protein? We constructed a derivative of $N_{ub}$-Sec62p in which the C-terminal 60 residues of Sec62p were replaced by the DHFR-ha (Dha) moiety, yielding Sec62(ΔC60)-Dha. A similar Sec62p-invertase fusion was described by Deshaies and Schekman (1990) and shown to be nonfunctional. As expected, neither Sec62(ΔC60)-Dha nor $N_{ub}$-Sec62(ΔC60)-Dha complemented the ts phenotype of RSY529 cells (Figure 3B).

The Ub-reconstitution activity of $N_{ub}$-Sec62(ΔC60)-Dha in the presence of either $Mf\alpha_{37}$-$C_{ub}$-Dha or Tpi1-$C_{ub}$-Dha (Figure 6, A and B) was compared with the activity of $N_{ub}$-Sec62p and $N_{ub}$-Sec61p in the presence of the same $C_{ub}$-containing fusions. Remarkably, no cleavage of $Mf\alpha_{37}$-$C_{ub}$-Dha was observed in the presence of $N_{ub}$-Sec62(ΔC60)-Dha, whereas the cytosolic Tpi1-$C_{ub}$-Dha was cleaved. This result (Figure 6, A and B) indicated that the concentration and accessibility of $N_{ub}$ were comparable for the functionally inactive $N_{ub}$-Sec62(ΔC60)-Dha and the functionally active $N_{ub}$-Sec62p. In these experiments, $N_{ub}$-Sec62(ΔC60)-Dha, which could be detected with the anti-ha antibody (Figure 6, A and B), was expressed from the uninduced $P_{CUP1}$ promoter. Strikingly, even overexpression of $N_{ub}$-Sec62(ΔC60)-Dha, from the copper-induced $P_{CUP1}$, did not result in a significant cleavage of $Mf\alpha_{37}$-$C_{ub}$-Dha (our unpublished results). These control experiments with the inactive derivative of Sec62p indicated that the proximity signal in the split-Ub assay with Sec62p requires the functional activity of Sec62p.

### *Using Ura3p Reporter to Detect the In Vivo Proximity between Sec62p and Signal Sequences*

The DHFR-ha (Dha) reporter moiety of $Mf\alpha_{37}$-$C_{ub}$-Dha was replaced by *S. cerevisiae* Ura3p (orotidine-5′-phosphate decarboxylase), yielding $Mf\alpha_{37}$-$C_{ub}$-Ura3p. The use of cytosolic Ura3p as a reporter for translocation across membranes is well documented (Maarse *et al.*, 1992; Johnsson and Varshavsky, 1994b; Ng *et al.*, 1996). The high sensitivity of Ura3p-based assays (cells become $Ura^+$ if a threshold amount of Ura3p is present in the cytosol) allowed us to express the $N_{ub}$ and $C_{ub}$ fusions from the uninduced $P_{CUP1}$ promoter. Since the efficient translocation of $Mf\alpha_{37}$-$C_{ub}$-Ura3p sequesters the Ura3p activity in the ER, a *ura3*Δ strain of *S. cerevisiae* that expressed $Mf\alpha_{37}$-$C_{ub}$-Ura3p remained $Ura^-$ (Figure 7A). $N_{ub}$-Sec62p, which, as shown above, is close to the nascent chain of $Mf\alpha_{37}$-$C_{ub}$-Dha during its translocation, induced enough cleavage of $Mf\alpha_{37}$-$C_{ub}$-Ura3p at the $C_{ub}$–Ura3p junction to render cells $Ura^+$ (Figure 7B). Cells were transformed with either $N_{ub}$-Sec62p, $N_{ub}$-Sec61p, $N_{ub}$-Sec62(ΔC60)-Dha, or $N_{ub}$-Bos1p to compare relative proximities of these $N_{ub}$-containing proteins to $C_{ub}$ fusions bearing the Ura3p reporter moiety and either

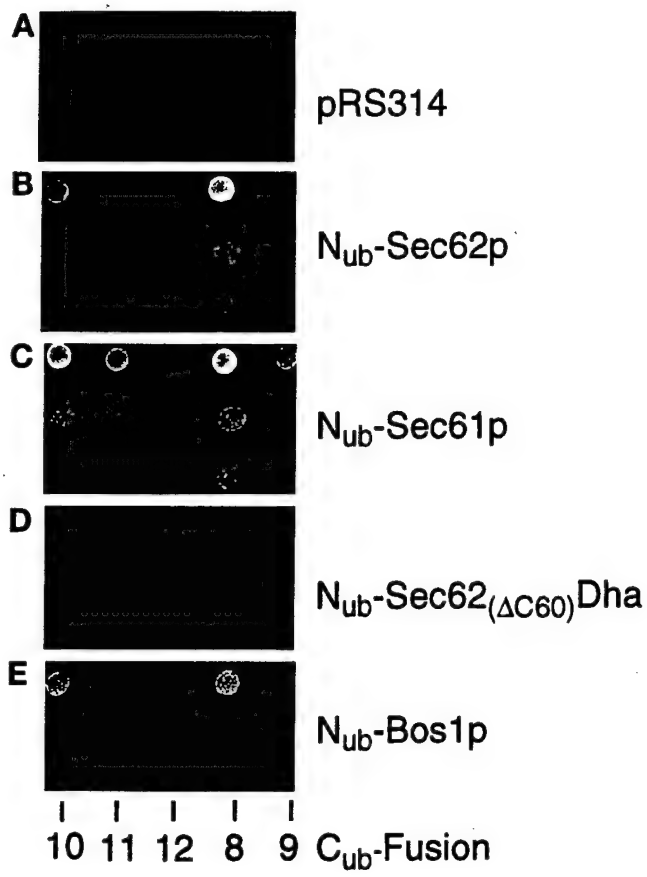



**Figure 7.** The use of a metabolic marker to assess the proximity between a component of the translocation machinery and a translocated protein. *S. cerevisiae* expressing the $C_{ub}$ fusions 8–12 (Figure 2) that contained Ura3p instead of Dha (see the main text) were transformed with the vector (A) or plasmids expressing $N_{ub}$-Sec62p (B), $N_{ub}$-Sec61p (C), $N_{ub}$-Sec62(ΔC60)Dha (D), and $N_{ub}$-Bos1p (E). Cells were grown in a liquid uracil-containing SD medium, and ~$10^5$, $10^3$, and $10^2$ cells were spotted onto uracil-lacking SD medium. Plates were examined after 18 h at 30°C.

the invertase-derived or the α-factor-derived signal sequence (Figure 7, B–E). The cells were spotted on plates lacking uracil and incubated at 30°C for 18 h. The growth patterns of strains that expressed different combinations of $N_{ub}$- and $C_{ub}$-containing fusions confirmed the results of analyses with analogous (but more highly expressed) Dha-based constructs.

In particular, the interaction of Sec62p with the signal sequence of prepro-α-factor was stronger than with the signal sequence of invertase. This proximity was not detectable when the distance between a signal sequence and the $C_{ub}$ moiety of a fusion was increased (Figure 7B). Sec61p appears to be equally close to both of the signal sequences tested. Again, the proximity signal was gradually lost when the distance between the signal sequence and the $C_{ub}$ moiety was increased (Figure 7C). Cells acquired a weak $Ura^+$ phenotype in the presence of $N_{ub}$-Bos1p and the $C_{ub}$–Ura3p fusions 8 and 10 (Figure 7E). If used as a reference to discriminate between specific and nonspecific signals in this assay, Sec62p, under these conditions, appears to interact only with the α-factor signal sequence. No interaction with any of the tested $C_{ub}$ constructs was detectable with the functionally inactive $N_{ub}$-Sec62(ΔC60)-Dha (Figure 7D).

## DISCUSSION

The new application of the split-Ub technique (Johnsson and Varshavsky, 1994a, 1997) described in the present work introduces a tool for the analysis of transient (short-lived) protein interactions in living cells. A split-Ub assay involves the tagging of two (presumably) interacting proteins with the N- and C-terminal halves of Ub, $N_{ub}$ and $C_{ub}$, and monitoring, in a variety of ways, the release of the reporter protein fused to the C terminus of $C_{ub}$. The reporter release, through the cleavage by Ub-specific processing proteases (UBPs), takes place in the cytosol if the two halves of Ub interact in vivo to form a quasi-native Ub moiety upstream of the reporter (Figure 1). Among the advantages of this method are its applicability either in living cells or in vitro and its sensitivity to kinetic aspects of a protein interaction.

In the present work, we applied the split-Ub technique to the problem of protein translocation across membranes. We showed that Sec62p of *S. cerevisiae* is spatially close to the signal sequence of the nascent α-factor polypeptide in vivo. This proximity is confined to the nascent polypeptide chain immediately following the signal sequence. In addition, the extent of proximity depends on the nature of the signal sequence. Specifically, $C_{ub}$-containing test proteins that bore the signal sequence of invertase resulted in a much lower Ub reconstitution with $N_{ub}$-Sec62p than the same $C_{ub}$-containing proteins bearing the signal sequence of α-factor. An inactive derivative of Sec62p failed to interact with signal sequences in the split-Ub assay. Taken together, these findings are the first in vivo evidence that *S. cerevisiae* Sec62p, a component of the ER translocation machinery, is a part of a signal sequence-binding complex.

### *In Vivo Proximity between Sec62p and the Signal Sequence of α-Factor*

We have previously shown that a region of the nascent polypeptide chain that lies close to the signal sequence of invertase or the prepro-α-factor is briefly exposed to the cytosol before its translocation into the ER (Johnsson and Varshavsky, 1994b). This feature of translocation enabled us, in the present work, to use the split-Ub assay for monitoring the proximity between a secretory protein and components of the translocation machinery. The $C_{ub}$ moiety was placed 37 residues

downstream from the signal sequence of the α-factor precursor, and the N terminus of Sec62p (see INTRODUCTION) was extended with $N_{ub}$. Using this version of the split-Ub assay, we could demonstrate that Sec62p is close to the nascent chain of the α-factor during its translocation. By moving the $C_{ub}$ moiety farther downstream from the signal sequence of the α-factor precursor, we obtained "snapshots" of the relative proximity between the nascent polypeptide chain and Sec62p in vivo. The proximity thus detected was considerably reduced once the spacer sequence between the signal sequence of the α-factor precursor and $C_{ub}$ was increased from 37 to 65 residues (Figures 5B and 7B). The data strongly suggest that the access of Sec62p to the nascent chain of α-factor is confined to a region of the nascent chain that is very close to the signal sequence. Similar results were obtained with the signal sequence of invertase as well. This property of Sec62p is the expected feature of a component of a signal sequence receptor.

Our interpretation, supported by several control experiments, is in agreement with the results produced by cross-linking and binding studies in cell-free systems (Müsch *et al.*, 1992; Lyman and Schekman, 1997; Matlack *et al.*, 1997). Sec62p could be cross-linked to the α-factor precursor in vitro, but only when ATP was omitted and the initiation of translocation of α-factor was halted. Upon the addition of ATP, the translocation resumed and cross-linking was no longer possible (Müsch *et al.*, 1992; Lyman and Schekman, 1997). The cross-linking between Sec62p and the nascent polypeptide chain was not observed when the translocating chain was halted in the ER channel (Müsch *et al.*, 1992; Sanders *et al.*, 1992). It was therefore assumed that Sec62p is not a part of the channel and that it functions in the early steps of substrate recognition and initiation of translocation. The split-Ub assay, in its current form, depends on both halves of Ub being in the cell's cytosol. Therefore, the absence of the diagnostic cleavage (Figures 5B, 5C, and 7) when the $C_{ub}$ moiety was placed farther downstream from the signal sequence (see RESULTS), while consistent with the absence of interactions between Sec62p and the nascent chain after the initiation of its translocation, does not address this issue directly. The $C_{ub}$ moiety that emerges from the ribosome after it has docked at the ER channel is not accessible to the cytosolic $N_{ub}$ moiety even if an $N_{ub}$-linked protein is spatially close to the translocation pore. This also explains the inability of $N_{ub}$-Sec61p to induce the cleavage of $C_{ub}$-containing translocation substrates bearing long spacer sequences between the signal sequence and $C_{ub}$, although the in vitro cross-linking studies have shown Sec61p to be in constant contact with the translocating polypeptide (Mothes *et al.*, 1994) (Figure 7).

The proximity between Sec62p (or Sec61p) and a translocating polypeptide is short lived. The rapid transfer of the $C_{ub}$ moiety into the lumen of the ER was shown to either prevent or strongly inhibit its interaction with the $N_{ub}$ moiety of $N_{ub}$-Sec62p and $N_{ub}$-Sec61p (Figures 3A and 6C). In these experiments, the $N_{ub}$ moieties bore either the glycine ($N_{ug}$) or the alanine ($N_{ua}$) residue at position 13 of $N_{ub}$. These modifications decrease the affinity between the two halves of Ub (see INTRODUCTION), thereby making the reconstitution of a quasi-native Ub moiety more dependent on the stability (half-life) of interactions between the proteins linked to $N_{ub}$ and $C_{ub}$. These assays clearly distinguished the Sec62p–Sec61p signal sequence interactions from those that underlie the better understood, longer-lived protein complexes. For example, when linked to homodimerizing leucine zippers, the $N_{ua}$ moiety, and even the $N_{ug}$ moiety, is sufficient for reconstitution of the Ub (Johnsson and Varshavsky, 1994a).

### *Is Sec62p Part of a Signal Sequence Receptor?*

The split-Ub assay measures the concentrations of the protein-coupled $N_{ub}$ and $C_{ub}$ moieties in the immediate vicinity of each other. Therefore, a positive result of a split-Ub assay signifies a spatial proximity between the two proteins but cannot, by itself, prove their physical interaction or address the functional significance of this proximity. $N_{ub}$-Sec62(ΔC60)Dha is functionally inactive and was shown to be not close to the translocating $Mf\alpha_{37}$-$C_{ub}$-Dha (Figures 6 and 7). Since the sequence between $N_{ub}$ and the first membrane-spanning region of Sec62 was retained in the C-terminally truncated Sec62(ΔC60)Dha, the distance between $N_{ub}$ and the ER membrane was, most probably, not altered relative to wild-type Sec62p. The lack of significant cleavage of $Mf\alpha_{37}$-$C_{ub}$-Dha in the presence of $N_{ub}$-Sec62(ΔC60)Dha must therefore result from the increased distance between Sec62(ΔC60)p and the translocating polypeptide chain. The C-terminal domain of intact Sec62p may contact other components of the translocation complex; alternatively, it may contribute to a binding site for the signal sequence or the nascent chain. These and related uncertainties notwithstanding, our results (Figures 6 and 7) provide the first in vivo evidence to support the view that Sec62p is part of a signal sequence-binding complex.

### *Sec62p Discriminates between Different Signal Sequences*

The split-Ub assay has made it possible to show that Sec62p discriminates, in living cells, between two distinct signal sequences in otherwise identical fusion proteins. In the presence of $N_{ub}$-Sec62p, more of the free Dha reporter protein was produced in vivo from $Mf\alpha_{37}$-$C_{ub}$-Dha than from $Suc2_{23}$-$C_{ub}$-Dha (Figures 5 and 6). This selectivity is a property of Sec62p and not

a feature of the assay used, since approximately equal amounts of the cleaved reporter were produced with different signal sequences if the $N_{ub}$ moiety was present as the $N_{ub}$–Sec61p fusion (Figures 6 and 7). This result confirmed, in vivo, that different targeting pathways converge at the Sec61-containing complex to initiate translocation. The existence of at least two different targeting pathways to the translocation pore was suggested by Walter and co-workers on the basis of the properties of yeast mutants that lacked either SRP or its receptor (Hann and Walter, 1991; Ogg *et al.*, 1992). One possibility is that the targeting via SRP operates cotranslationally, whereas the targeting via the Sec62/63 complex is predominantly posttranslational. The bifurcation between the cotranslational and posttranslational targeting is expected to be stochastic for many translocated proteins. Nonetheless, certain signal sequences do prefer SRP, while some of the other signal sequences are targeted by the Sec62/Sec63 complex (Ng *et al.*, 1996). Genetic studies have shown that the translocation of invertase continues in the presence of mutations in either the SRP or the Sec62/63 complex (Deshaies and Schekman, 1989; Ogg *et al.*, 1992). However, the kinetics and efficiency of invertase translocation are altered in the absence of SRP (Hann and Walter, 1991; Johnsson and Varshavsky, 1994b).

To explain the different efficiencies of cleavage of the two signal sequence-bearing $C_{ub}$ fusions, we propose the following model. The signal sequence of invertase is recognized primarily by SRP and then transferred to the trimeric Sec61p complex, which completes the protein's translocation across the ER membrane. Under conditions that result in a shortage of SRP or a competition among different signal sequences, the Sec62p/Sec63p-containing complex, being a part of the alternative targeting pathway, would recognize an increasing fraction of invertase. This would explain why specific interactions between the invertase signal sequence-bearing proteins and Sec62p can only be observed for the more highly expressed $C_{ub}$–Dha fusions (Figure 5C). By contrast, the targeting of proteins bearing the signal sequence of the α-factor precursor is mediated, *in vivo*, predominantly by the Sec62p/Sec63p-containing complex. This would account for the observed close proximity of these test proteins to Sec62p (Figures 5–7). If so, the split-Ub technique makes it possible to estimate the flux of two different secretory proteins through the targeting pathways to the ER membrane without the necessity of deleting or otherwise inactivating specific components of the targeting complex.

### *Further Applications of the Split-Ub Technique*

The split-Ub sensor should also be applicable to other settings that involve short-lived protein interactions that occur in the cytosol and are freely accessible to the Ub-specific proteases. The advantage of using this method for the analysis of protein translocation stems, in part, from the fast and irreversible removal of the translocated chain from the location (cytosol) where the $N_{ub}/C_{ub}$ interaction is monitored. Similar situations are expected for the translocation of proteins into other organelles such as the mitochondrion, the nucleus, and the peroxisome.

The demonstration, in the present work, that Ura3p can serve as a reporter in a split-Ub assay (Figure 7) opens the way to genetic screens based on this assay. For example, introducing a DNA library consisting of random $N_{ub}$–gene fusions into a strain expressing a signal sequence-bearing $C_{ub}$-Ura3p protein should allow the identification of genes involved in targeting or translocation by enabling the cells to form colonies on media lacking uracil. This selection for protein ligands that interact transiently in the vicinity of a membrane complements the recent split-Ub–based screen for ligands that form relatively stable complexes (Stagljar *et al.*, 1998).

## ACKNOWLEDGMENTS

We thank Ray Deshaies, Jürgen Dohmen, Nicole Lewke, Randy Schekman, and Sandra Wittke for the gifts of yeast strains and plasmids. M.D. and N.J. thank Silke Müller for excellent technical assistance. This work was supported by a grant to N.J. from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (0311107), and a grant to A.V. from the National Institutes of Health (GM-31530).

## REFERENCES

Allison, D.S., and Young, E.T. (1988). Single-amino-acid substitutions within the signal sequence of yeast prepro-α-factor affect membrane translocation. Mol. Cell. Biol. *8*, 1915–1922.

Aronheim, A., Zandi, E., Hennemann, H., Elledge, S.J., and Karin, M. (1997). Isolation of an AP-1 repressor by a novel method for detecting protein- protein interactions. Mol. Cell. Biol. *17*, 3094–3102.

Banfield, D.K., Lewis, M.J., Rabouille, C., Warren, G., and Pelham, H.R. (1994). Localization of Sed5, a putative vesicle targeting molecule, to the *cis*- Golgi network involves both its transmembrane and cytoplasmic domains. J. Cell Biol. *127*, 357–371.

Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G., and Frank, J. (1997). Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. Science *278*, 2123–2126.

Biederer, T., Volkwein, C., and Sommer, T. (1996). Degradation of subunits of the Sec61p complex, an integral component of the ER membrane, by the ubiquitin-proteasome pathway. EMBO J. *15*, 2069–2076.

Bird, P., Gething, M.J., and Sambrook, J. (1987). Translocation in yeast and mammalian cells: not all signal sequences are functionally equivalent. J. Cell Biol. *105*, 2905–2914.

Brodsky, J.L., and Schekman, R. (1993). A Sec63p-BiP complex from yeast is required for protein translocation in a reconstituted proteoliposome. J. Cell Biol. *123*, 1355–1363.

Connolly, T., Collins, P., and Gilmore, R. (1989). Access of proteinase K to partially translocated nascent polypeptides in intact and detergent-solubilized membranes. J. Cell Biol. *108*, 299–307.

Crowley, K.S., Liao, S., Worrell, V.E., Reinhart, G.D., and Johnson, A.E. (1994). Secretory proteins move through the endoplasmic reticulum membrane via an aqueous, gated pore. Cell *78*, 461–471.

Crowley, K.S., Reinhart, G.D., and Johnson, A.E. (1993). The signal sequence moves through a ribosomal tunnel into a noncytoplasmic aqueous environment at the ER membrane early in translocation. Cell *73*, 1101–1115.

Deshaies, R.J., Sanders, S.L., Feldheim, D.A., and Schekman, R. (1991). Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. Nature *349*, 806–808.

Deshaies, R.J., and Schekman, R. (1989). SEC62 encodes a putative membrane protein required for protein translocation into the yeast endoplasmic reticulum. J. Cell Biol. *109*, 2653–2664.

Deshaies, R.J., and Schekman, R. (1990). Structural and functional dissection of Sec62p, a membrane-bound component of the yeast endoplasmic reticulum protein import machinery. Mol. Cell. Biol. *10*, 6024–6035.

Dohmen, R.J., Stappen, R., McGrath, J.P., Forrova, H., Kolarov, J., Goffeau, A., and Varshavsky, A. (1995). An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. J. Biol. Chem. *270*, 18099–18109.

Feldheim, D., and Schekman, R. (1994). Sec72p contributes to the selective recognition of signal peptides by the secretory polypeptide translocation complex. J. Cell Biol. *126*, 935–943.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245–246.

Gilmore, R., and Blobel, G. (1985). Translocation of secretory proteins across the microsomal membrane occurs through an environment accessible to aqueous perturbants. Cell *42*, 497–505.

Görlich, D., Prehn, S., Hartmann, E., Kalies, K.U., and Rapoport, T.A. (1992). A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. Cell *71*, 489–503.

Hanein, D., Matlack, K.E., Jungnickel, B., Plath, K., Kalies, K.U., Miller, K.R., Rapoport, T.A., and Akey, C.W. (1996). Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. Cell *87*, 721–732.

Hann, B.C., and Walter, P. (1991). The signal recognition particle in *S. cerevisiae*. Cell *67*, 131–144.

Hiller, M.M., Finger, A., Schweiger, M., and Wolf, D.H. (1996). ER degradation of a misfolded luminal protein by the cytosolic ubiquitin-proteasome pathway. Science *273*, 1725–1728.

Johnsson, N., and Varshavsky, A. (1994a). Split ubiquitin as a sensor of protein interactions in vivo. Proc. Natl. Acad. Sci. USA *91*, 10340–10344.

Johnsson, N., and Varshavsky, A. (1994b). Ubiquitin-assisted dissection of protein transport across membranes. EMBO J. *13*, 2686–2698.

Johnsson, N., and Varshavsky, A. (1997). Split ubiquitin: a sensor of protein interactions in vivo. In: The Yeast Two Hybrid System, ed. P.L. Bartel and S. Fields, Oxford, United Kingdom: Oxford University Press, 316–332.

Jungnickel, B., and Rapoport, T.A. (1995). A posttargeting signal sequence recognition event in the endoplasmic reticulum membrane. Cell *82*, 261–270.

Krieg, U.C., Walter, P., and Johnson, A.E. (1986). Photocrosslinking of the signal sequence of nascent preprolactin to the 54-kilodalton polypeptide of the signal recognition particle. Proc. Natl. Acad. Sci. USA *83*, 8604–8608.

Kurzchalia, T.V., Wiedmann, M., Girshovich, A.S., Bochkareva, E.S., Bielka, H., and Rapoport, T.A. (1986). The signal sequence of nascent preprolactin interacts with the 54K polypeptide of the signal recognition particle. Nature *320*, 634–636.

Lyman, S.K., and Schekman, R. (1997). Binding of secretory precursor polypeptides to a translocon subcomplex is regulated by BiP. Cell *88*, 85–96.

Maarse, A.C., Blom, J., Grivell, L.A., and Meijer, M. (1992). MPI1, an essential gene encoding a mitochondrial membrane protein, is possibly involved in protein import into yeast mitochondria. EMBO J. *11*, 3619–3628.

Matlack, K.E., Plath, K., Misselwitz, B., and Rapoport, T.A. (1997). Protein transport by purified yeast Sec complex and Kar2p without membranes. Science *277*, 938–941.

Miyawaki, A., Llopis, J., Heim, R., McCaffery, J.M., Adams, J.A., Ikura, M., and Tsien, R.Y. (1997). Fluorescent indicators for $Ca^{2+}$ based on green fluorescent proteins and calmodulin. Nature *388*, 882–887.

Mothes, W., Prehn, S., and Rapoport, T.A. (1994). Systematic probing of the environment of a translocating secretory protein during translocation through the ER membrane. EMBO J. *13*, 3973–3982.

Müsch, A., Wiedmann, M., and Rapoport, T.A. (1992). Yeast Sec proteins interact with polypeptides traversing the endoplasmic reticulum membrane. Cell *69*, 343–352.

Nakano, A., Brada, D., and Schekman, R. (1988). A membrane glycoprotein, Sec12p, required for protein transport from the endoplasmic reticulum to the Golgi apparatus in yeast. J. Cell Biol. *107*, 851–863.

Ng, D.T., Brown, J.D., and Walter, P. (1996). Signal sequences specify the targeting route to the endoplasmic reticulum membrane. J. Cell Biol. *134*, 269–278.

Ogg, S.C., Poritz, M.A., and Walter, P. (1992). Signal recognition particle receptor is important for cell growth and protein secretion in *Saccharomyces cerevisiae*. Mol. Biol. Cell *3*, 895–911.

Orlean, P., Kuranda, M.J., and Albright, C.F. (1991). Analysis of glycoproteins from *Saccharomyces cerevisiae*. Methods Enzymol. *194*, 682–697.

Panzner, S., Dreier, L., Hartmann, E., Kostka, S., and Rapoport. T.A. (1995). Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. Cell *81*, 561–570.

Plemper, R.K., Böhmler, S., Bordallo, J., Sommer, T., and Wolf, D.H. (1997). Mutant analysis links the translocon and BiP to retrograde protein transport for ER degradation. Nature *388*, 891–895.

Rapoport, T.A., Jungnickel, B., and Kutay, U. (1996). Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. Annu. Rev. Biochem. *65*, 271–303.

Rossi, F., Charlton, C.A., and Blau, H.M. (1997). Monitoring protein-protein interactions in intact eukaryotic cells by $\beta$-galactosidase complementation. Proc. Natl. Acad. Sci. USA *94*, 8405–8410.

Rothblatt, J.A., Deshaies, R.J., Sanders, S.L., Daum, G., and Schekman, R. (1989). Multiple genes are required for proper insertion of secretory proteins into the endoplasmic reticulum in yeast. J. Cell Biol. *109*, 2641–2652.

Sanders, S.L., Whitfield, K.M., Vogel, J.P., Rose, M.D., and Schekman, R.W. (1992). Sec61p and BiP directly facilitate polypeptide translocation into the ER. Cell *69*, 353–365.

Sapperstein, S., Berkower, C., and Michaelis, S. (1994). Nucleotide sequence of the yeast STE14 gene, which encodes farnesylcysteine

carboxyl methyltransferase, and demonstration of its essential role in a-factor export. Mol. Cell. Biol. *14*, 1438–1449.

Shim, J., Newman, A.P., and Ferro-Novick, S. (1991). The BOS1 gene encodes an essential 27-kDa putative membrane protein that is required for vesicular transport from the ER to the Golgi complex in yeast. J. Cell Biol. *113*, 55–64.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics *122*, 19–27.

Simon, S.M., and Blobel, G. (1991). A protein-conducting channel in the endoplasmic reticulum. Cell *65*, 371–380.

Stagljar, I., Korostensky, C., Johnsson, N., and te Heesen, S. (1998). A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. Proc. Natl. Acad. Sci. USA *95*, 5187–5192.

Walter, P., and Johnson, A.E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. Annu. Rev. Cell Biol. *10*, 87–119.

Walter, P., Ibrahimi I., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum I. Signal recognition protein (SRP) binds to in vitro assembled polysomes synthesizing secretory protein. J. Cell Biol. *91*, 545–550.

Wiertz, E.J., Tortorella, D., Bogyo, M., Yu, J., Mothes, W., Jones, T.R., Rapoport, T.A., and Ploegh, H.L. (1996). Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction. Nature *384*, 432–438.

Wilkinson, B.M., Critchley, A.J., and Stirling, C.J. (1996). Determination of the transmembrane topology of yeast Sec61p, an essential component of the endoplasmic reticulum translocation complex. J. Biol. Chem. *271*, 25590–25597.

Cell, Vol. 92, 489–499, February 20, 1998, Copyright ©1998 by Cell Press

# Ump1p Is Required for Proper Maturation of the 20S Proteasome and Becomes Its Substrate upon Completion of the Assembly

Paula C. Ramos,* Jörg Höckendorff,*
Erica S. Johnson,†‡ Alexander Varshavsky,†
and R. Jürgen Dohmen*§
*Biotechnologisches Zentrallabor
Institut für Mikrobiologie
Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1, Geb. 25.02
40225 Düsseldorf
Germany
†Division of Biology
California Institute of Technology
Pasadena, California 91125
‡Laboratory of Cell Biology
The Rockefeller University
New York, New York 10021

## Summary

**We report the discovery of a short-lived chaperone that is required for the correct maturation of the eukaryotic 20S proteasome and is destroyed at a specific stage of the assembly process. The *S. cerevisiae* Ump1p protein is a component of proteasome precursor complexes containing unprocessed β subunits but is not detected in the mature 20S proteasome. Upon the association of two precursor complexes, Ump1p is encased and is rapidly degraded after the proteolytic sites in the interior of the nascent proteasome are activated. Cells lacking Ump1p exhibit a lack of coordination between the processing of β subunits and proteasome assembly, resulting in functionally impaired proteasomes. We also show that the propeptide of the Pre2p/Doa3p β subunit is required for Ump1p's function in proteasome maturation.**

## Introduction

Ubiquitin (Ub)–dependent proteolysis underlies the bulk of nonlysosomal protein degradation in eukaryotic cells (reviewed by Hochstrasser, 1996; Varshavsky, 1997). Naturally short-lived as well as damaged or otherwise abnormal proteins are recognized by the Ub system and are marked for degradation by the attachment of multi-Ub chains. Ubiquitylated proteins are degraded by the 26S proteasome, an ~2000 kDa, multisubunit, ATP-dependent protease that consists of the 19S complex, which is required specifically for the degradation of ubiquitylated proteins, and an ~700 kDa complex, called the 20S proteasome, which is the ATP-independent catalytic core of the 26S proteasome (reviewed by Peters, 1994; Coux et al., 1996; Lupas et al. 1997).

The 20S proteasome is universal among eukaryotes; its structural homologs have also been found in archaeons and eubacteria (reviewed by Coux et al., 1996). High-resolution crystal structures have been reported for the 20S proteasomes of the archaeon *Thermoplasma acidophilum* and the eukaryote *Saccharomyces cerevisiae* (Löwe et al., 1995; Groll et al., 1997). The *Thermoplasma* proteasome contains two types of subunits, α and β, which form a hollow cylinder composed of four heptameric rings in the configuration $\alpha_7\beta_7\beta_7\alpha_7$. The yeast 20S proteasome has a similar structure but contains 14 distinct subunits, seven of the α type and seven of the β type. The N-terminal threonines of the β subunits of the *Thermoplasma* proteasome act as nucleophiles in catalyzing the hydrolysis of peptide bonds of polypeptide substrates (Seemüller et al., 1995). Lys-33 and Glu-17 of the β subunit also play a role in catalyzing the cleavage of peptide bonds. The 14 identical β subunits of the *Thermoplasma* proteasome form 14 identical active sites, which catalyze the cleavage of substrates after hydrophobic amino acid residues (chymotrypsin-like active sites). These sites are located on the inner surface of a central chamber formed by the two rings of β subunits (Löwe et al., 1995; Seemüller et al., 1995). β subunits are synthesized as inactive precursors containing a propeptide that is thought to be cleaved off autocatalytically, yielding the mature β subunit bearing N-terminal Thr. The propeptides of β subunits are not required for the in vitro assembly of the *Thermoplasma* proteasome (Seemüller et al., 1996).

In eukaryotic proteasomes, only 3 of the 7 distinct β subunits contain the three conserved residues required for activity of the *Thermoplasma* proteasome. Genetic and structural data suggest that these three subunits provide the active-site nucleophiles for the three distinct catalytic activities of eukaryotic proteasomes, namely the "chymotrypsin-like" (see above), the "trypsin-like" (cleavage after basic residues), and the "peptidylglutamyl peptide hydrolyzing (PGPH)" activity (cleavage after acidic residues). In *S. cerevisiae*, these three β subunits are Pre2p/Doa3p, Pup1p, and Pre3p (Heinemeyer et al., 1993; Chen and Hochstrasser, 1996; Arendt and Hochstrasser, 1997; Groll et al., 1997; Heinemeyer et al., 1997). In the human 20S proteasome, the related proteins are MB1, δ, and Z; during the immune response, these subunits can be replaced by their respective homologs LMP7, LMP2, and MECL-1/LMP10 (reviewed by Coux et al., 1996). In the case of LMP2 and LMP7, it has been demonstrated that these replacements are functionally relevant in altering the specificity of antigen presentation by the MHC class I pathway. The catalytically active β subunits of the eukaryotic proteasome are synthesized with propeptides, similarly to the β subunit in *Thermoplasma* (Schmidtke et al., 1997).

In the crystal structure of the *S. cerevisiae* 20S proteasome, the opening to the proteasome's interior, formed by the outer ring of the α subunits, is not large enough to admit even an unfolded polypeptide chain, let alone a folded protein (Groll et al., 1997). The degradation of larger substrates requires the 19S complexes, which attach at both sides of the 20S proteasome, yielding the 26S proteasome. The 19S complexes contain subunits that bind multi-Ub chains, at least one Ub-specific isopeptidase that disassembles these chains, and several

§To whom correspondence should be addressed.

ATPases that are thought to be involved in perturbing the substrate conformationally and guiding it to the interior of the 20S proteasome (Deveraux et al., 1994; Hochstrasser et al., 1995; Jentsch and Schlenker, 1995; Coux et al., 1996).

Studies with mammalian cells have shown that the 20S proteasome is assembled through a 15–16S intermediate, apparently a half-proteasome. This intermediate contains all 14 α and β subunits, some of which are in the precursor (propeptide-bearing) form, and several uncharacterized polypeptides as well (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidtke et al., 1997). Studies on the yeast 20S proteasome have demonstrated that processing of proPre2p (identical to proDoa3p) is coupled to formation of the 20S proteasome from two half-proteasome precursors (Chen and Hochstrasser, 1996). Formation of the active site capable of autocatalytic processing of proPre2p depends on the juxtaposition of proPre2p and Pre1p on the opposite sides of the two halves of the proteasome. This mechanism is thought to prevent activation of proteolytic sites before the central hydrolytic chamber has been sealed off from the cytosol through association of the two halves of the proteasome (Chen and Hochstrasser, 1996). The Pre2p propeptide is essential for the formation of functional proteasomes. Moreover, this propeptide can operate in *trans*, suggesting that it serves a chaperone-like function in proteasome maturation (Chen and Hochstrasser, 1996).

In the present work, we identify Ump1p, a novel protein, as a component of a precursor complex of the 20S proteasome. This precursor contains unprocessed β subunits. Upon formation of the 20S proteasome from two such precursors, the propeptides of β subunits are removed, a process that is accompanied by removal of Ump1p from the proteasome through Ump1p degradation, which requires the proteasome's proteolytic activity. In *ump1*Δ cells, which lack Ump1p, coordination of 20S complex formation and processing of β subunits is impaired, resulting in incompletely or (in the case of Pre2p) prematurely processed β subunits. These findings reveal Ump1p as a novel type of molecular chaperone, a short-lived maturation factor required for the efficient biogenesis of the 20S proteasome. We also describe genetic evidence that the propeptide of proPre2p is required for Ump1p-dependent proteasome maturation, and we present a model that accounts for some of the functions of Ump1p and propeptides in proteasome maturation.

## Results

### The *UMP1* Gene Is Required for Ubiquitin-Mediated Proteolysis

To identify the genes required for Ub-dependent proteolysis in *S. cerevisiae*, we screened for mutants defective in the degradation of test substrates. One such mutant, termed *ump1-1* (ub-mediated proteolysis), exhibited defects in the degradation of several normally short-lived proteins (see below). The complementing *UMP1* gene encoded a 148-residue (16.8 kDa) protein. Searches of the databases did not identify close homologs of Ump1p but did detect similarities to regions of several proteins. One intriguing similarity was between Ump1p and C-terminal regions of the protease inhibitor contrapsin (25% identity, 52% similarity) and related proteins (Figure 1A), consistent with the likely role of Ump1p as an inhibitor of premature proteolytic activation of the proteasome precursor complexes (see below).

Pulse-chase analyses in *ump1*Δ mutants revealed a strong stabilization of several normally short-lived test proteins, in particular an N-end rule substrate Arg–β-galactosidase (Arg–β-gal), which bears N-terminal Arg, a destabilizing residue (Varshavsky, 1996), Ub-Pro–β-gal (a substrate of the UFD pathway; Johnson et al., 1995; Varshavsky, 1997), and $\alpha 2_{deg1}$–β-gal (a substrate of the DOA pathway; Hochstrasser and Varshavsky, 1990; Hochstrasser et al., 1995) (Figure 1B). The recognition and ubiquitylation of these substrates involve different recognins (E3 proteins) and different Ub-conjugating (E2) enzymes. The pleiotropic character of the *ump1*Δ phenotype suggested that Ump1p functions downstream of the recognition and ubiquitylation components of the Ub system, perhaps at the step of proteolysis, or in regulating the supply of Ub.

### *ump1*Δ Mutants Are Hypersensitive to a Variety of Stresses, Accumulate Ub-Protein Conjugates, and Do Not Sporulate

*ump1*Δ mutants grew more slowly than congenic wild-type (wt) cells at 30°C or lower temperatures and were severely growth-impaired at higher temperatures (37°C) (Figure 1C and data not shown). In addition, they were hypersensitive to cadmium ions and to the arginine analog canavanine (Figure 1C). Similar phenotypes have been reported for mutants in genes encoding E2 (Ubc) enzymes or proteasome subunits (Heinemeyer et al., 1993; Jungmann et al., 1993). Comparisons of proteins from whole-cell extracts of wt and *ump1*Δ cells by immunoblotting with anti-Ub antibodies showed a dramatic accumulation of Ub-protein conjugates in the *ump1*Δ mutant. At the same time, the level of free Ub was reduced in *ump1*Δ cells (data not shown). Thus, the primary cause of the *ump1*Δ phenotype appears to be the impaired ability of *ump1*Δ cells to degrade Ub-protein conjugates. Homozygous *ump1*Δ diploids (strain JD61) were unable to sporulate (data not shown).

### Ump1p Is a Component of Proteasome Precursors

To investigate Ump1p biochemically, an epitope tag was linked to its C terminus. A single copy of the modified *UMP1* gene (*UMP1-ha*), expressed from its natural promoter and chromosomal location, restored wt growth rates. We analyzed whole-cell extracts of a strain expressing Ump1p-ha and Pre1p-ha (the latter an ha-tagged β subunit of the proteasome) by gel filtration on Superose-6. Ump1p-ha eluted in fractions 22 and 23, both of which also contained Pre1p-ha and other proteasomal subunits (Figures 2A, 2B, and 3A) but lacked the chymotrypsin-like activity of the mature 20S proteasome. The apparent size of the Ump1p-containing complex was 300–400 kDa (Figure 2A). 20S and 26S proteasomes, as well as free 19S caps, and 20S proteasomes with one attached 19S cap, eluted in earlier fractions

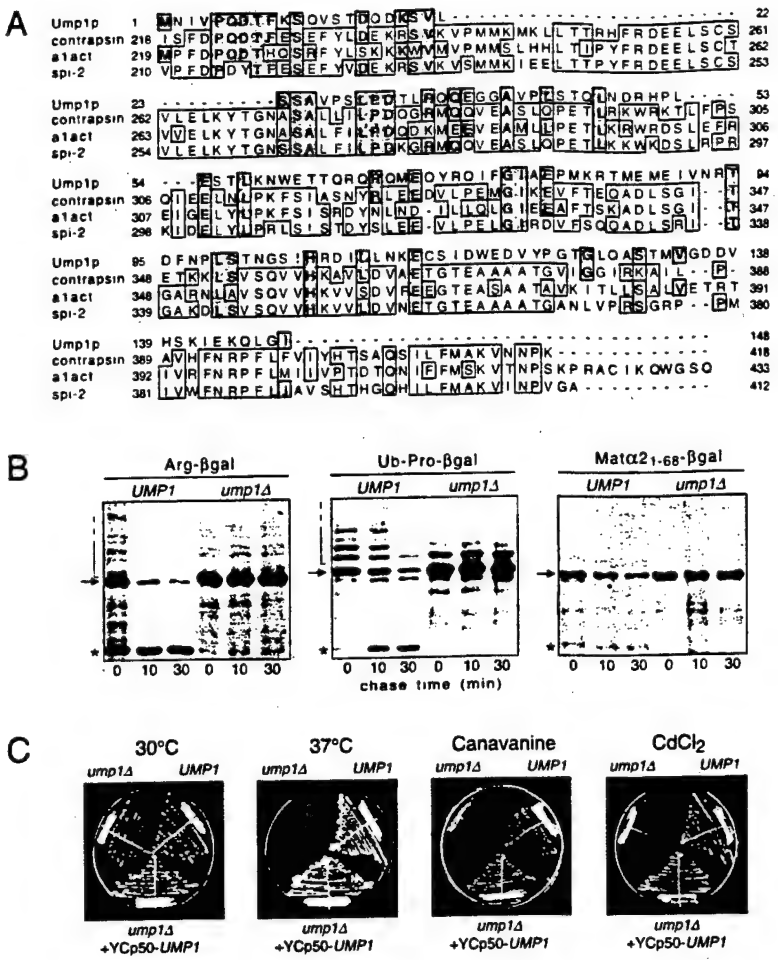



Figure 1. Ump1p Is Required for Ubiquitin-Mediated Proteolysis

(A) Sequence similarities between Ump1p and C-terminal parts of mouse contrapsin, human alpha1 antitrypsin inhibitor (a1act) and mouse spi-2 (PIR accession numbers, respectively, JX0129, A90475, and S31305). The sequences were aligned using the PileUp program (GCG package, version 7.2, Genetics Computer Group, Madison, WI). Gaps (indicated by hyphens) were used to maximize alignments. Residues identical between Ump1p and at least one of the other proteins are shaded in gray. Residues identical among at least two proteins other than Ump1p are boxed.

(B) Pulse-chase analysis comparing the metabolic stabilities of R–β-gal, Ub-P–β-gal, and MATα2$_{1-68}$–β-gal in wt (*UMP1*) and *ump1*Δ cells. The open-ended brackets denote the position of multiubiquitylated β-gal species. Asterisks denote an ~90 kDa β-gal cleavage product characteristic of short-lived β-gal derivatives (Dohmen et al., 1991).

(C) Growth of *ump1*Δ (JD59), *UMP1* (JD47-13C), and *ump1*Δ cells transformed with YCp50-*UMP1* (expressing wild-type *UMP1*). Cells were streaked on YPD plates and incubated for 2 days at 30°C (or at 37°C where indicated), with 0.8 μg/ml canavanine or 30 μm $CdCl_2$ where indicated.

(15–21), as determined by nondenaturing gel electrophoresis (Figure 2B and data not shown), by assays of the proteasome's proteolytic activity, and by immunological detection of Cim3p, a subunit specific for the 19S cap of the 26S proteasome (Ghislain et al., 1993).

Cochromatography of Ump1-ha and Pre1p-ha in fractions 22 and 23 (Figure 2A) suggested that Ump1p is a component of proteasome precursor complexes. Consistent with this possibility, the Ump1p-containing complex had a higher mobility than the 20S proteasome upon nondenaturing gel electrophoresis. A proteasome precursor with an electrophoretic mobility indistinguishable from that of the Ump1p-containing complex was also detected in extracts from a strain that expressed Pup1p-ha, another β subunit of the 20S proteasome (Figure 2B).

To produce independent evidence bearing on the nature of the Ump1p-containing complex, we constructed a strain that expressed both Ump1p-ha and doubly tagged Pre1p-Flag-His$_6$ (Pre1p-FH). Affinity chromatography on Ni-NTA-agarose and anti-Flag antibody resin was then used to purify complexes containing Pre1p-FH. Ump1p-ha cofractionated with Pre1p-FH in both affinity purification steps (Figures 2C and 2D), confirming that Ump1p is a component of a distinct proteasome-related complex. This complex sedimented at ~15S in sucrose gradients (data not shown). Taken together, this evidence strongly suggested that the Ump1p-containing complex is a precursor of the 20S proteasome that is similar to the 15–16S precursors observed in the maturation pathway of mammalian proteasomes (Frentzel et al., 1994; Yang et al., 1995).

### The Ump1p Proteasome Precursor Complex Contains Unprocessed β Subunits

Several subunits of the proteasome are synthesized as precursors containing N-terminal extensions (propeptides) that are absent from the mature proteasome. These precursors are detected in the mammalian 15–16S complex (Frentzel et al., 1994; Yang et al., 1995). To compare the Ump1p-containing complex with the mammalian 15–16S complex, we analyzed the former for the presence of propeptide-containing proteasomal subunits. These β subunits were modified by the addition of C-terminal ha tags. Strains expressing Pup1p-ha, Pre2p-ha, and Pre3p-ha grew at wt rates (data not shown). When extracts of these strains were fractionated by gel filtration on Superose-6, the fractions containing the Ump1p complex contained largely the β subunit precursors, proPup1-ha, proPre2p-ha, and proPre3p-ha (Figure 3A). The corresponding mature β subunits largely eluted in the earlier fractions that contained 20S and 26S proteasomes (Figure 3A). Subtle but reproducible differences

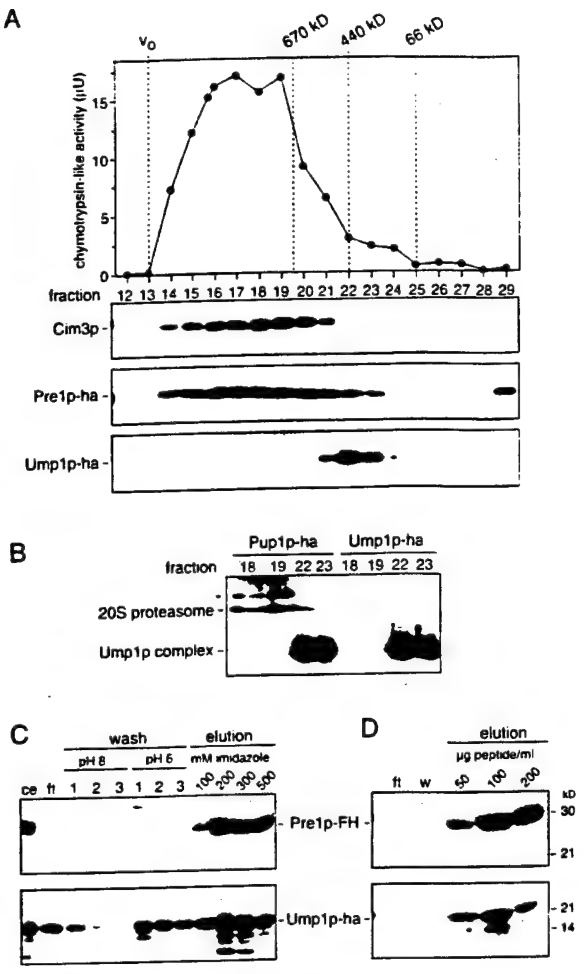



Figure 2. Ump1p Is a Component of Proteasome Precursor Complexes

(A) An extract from strain JD127 expressing Ump1p-ha and Pre1p-ha instead of wt versions of these proteins was fractionated by gel filtration on Superose-6. The upper panel shows the results of measurements of chymotrypsin-like activity in the relevant fractions (12–29). The last fraction, representing the void volume ($V_o$), and positions of the peaks of marker proteins are indicated by dashed lines. The same fractions were analyzed by SDS-PAGE and immunoblotting with anti-Cim3p and anti-ha antibodies, which detected Cim3p, Pre1p-ha, and Ump1p-ha, as indicated.

(B) Analysis of Superose-fractionated 20S proteasomes and precursor complexes by nondenaturing gel electrophoresis and immunoblotting with anti-ha antibody. The strains were JD139 and JD129 expressing, respectively, Pup1p-ha and Ump1p-ha. Asterisk, 20S proteasome with one attached 19S regulator cap. The 26S proteasome eluted in fractions 14–17 of Superose-6 (not shown).

(C) Affinity purification of Pre1p-FH on Ni-(NTA)-Sepharose and copurification of Ump1p-ha. Purification of proteins from an extract of strain JD126 expressing Ump1p-ha and Pre1p-FH was carried out as described in the Experimental Procedures. Different fractions were analyzed by SDS-PAGE and immunoblotting with anti-Flag (upper panel) and anti-ha antibodies (lower panel). Ce, crude extract; ft, flow-through. In contrast, no binding of Ump1p to the resin was observed when extracts from strain JD129 expressing untagged Pre1p were used (not shown).

(D) Affinity purification of Pre1p-FH on anti-Flag antibody agarose resin. Material that was eluted from the Ni-(NTA)-Sepharose column shown in panel C was subjected to a second affinity purification on anti-Flag antibody resin. Pre1p-FH was specifically eluted with increasing concentrations of the Flag peptide. W, wash.

were detected among the tagged β subunits with respect to the distribution of their precursor and processed forms in various fractions. For example, pro-Pup1p-ha was detectable only in fractions 22 and 23, and was absent from fraction 21. By contrast, proPre2p-ha was detectable in the larger complexes (down to fraction 19), while fractions 22 and 23 already contained some processed Pre3p-ha. These patterns suggested a defined order of processing events. One conclusion from these experiments is that the Ump1p-containing complex is a precursor of the 20S proteasome that contains unprocessed β subunits and is therefore proteolytically inactive (Figures 3A and 3C).

### Ump1p Is Required for Correct Proteasome Maturation

Next, we asked what effect the *ump1*Δ mutation has on maturation and activity of the proteasome. Specifically, the analyses described above were repeated with extracts obtained from *ump1*Δ cells. Figure 3C shows that there was a significant reduction in the three proteolytic activities of the proteasome in the fractions containing the 20S and 26S proteasomes (fractions 14–21). The reduction of the specific activity appears to be partially compensated by an increased expression of proteasomes (Figures 3A and 3B, data not shown) similar to that observed in cells expressing mutant β subunits (Arendt and Hochstrasser, 1997). In addition, a new (absent from wt cells) peak of chymotrypsin-like activity encompassing fractions 22 and 23 was detected in extracts from *ump1*Δ cells. These were the fractions that contained the Ump1p complex from wt extracts (see above and Figure 3A), suggesting that the absence of Ump1p from the proteasome precursor complex was the cause for premature activation of its chymotryptic activity in the *ump1*Δ mutant. To verify that the detected activity was actually of the ~15S precursor complex from *ump1*Δ cells, the fraction 22 samples from strains expressing either wt Pre2p or Pre2p-ha were incubated with anti-ha antibody and protein A Sepharose. This treatment did not deplete the chymotryptic activity from the sample containing untagged Pre2p but did deplete ~80% of the activity from an otherwise identical sample derived from *ump1*Δ cells expressing Pre2p-ha (data not shown). Thus, the chymotryptic activity of this fraction resided in the proteasome precursor complexes.

To follow the appearance of the precursor and mature forms of β subunits in proteasome precursors and mature proteasomes, we analyzed extracts from *ump1*Δ cells expressing ha-tagged versions of these subunits by immunoblotting with anti-ha antibody. The processing of the three analyzed β subunits, Pup1p-ha, Pre2p-ha, and Pre3p-ha, was strikingly different in *ump1*Δ and wt cells (Figure 3A). For all three subunits, a dramatic increase of their precursors was detected in the fractions (14–21) that contained the 20S and 26S forms of the proteasome. In addition, different processed variants, possibly representing processing intermediates of pro-Pre2p, could be detected in fractions 22 and 23, which contained the ~15S proteasome precursor, and in some of the proteasome-containing fractions as well (Figure 3A). In contrast, with the exception of Pre3p, almost no processed β subunits were present in the fractions

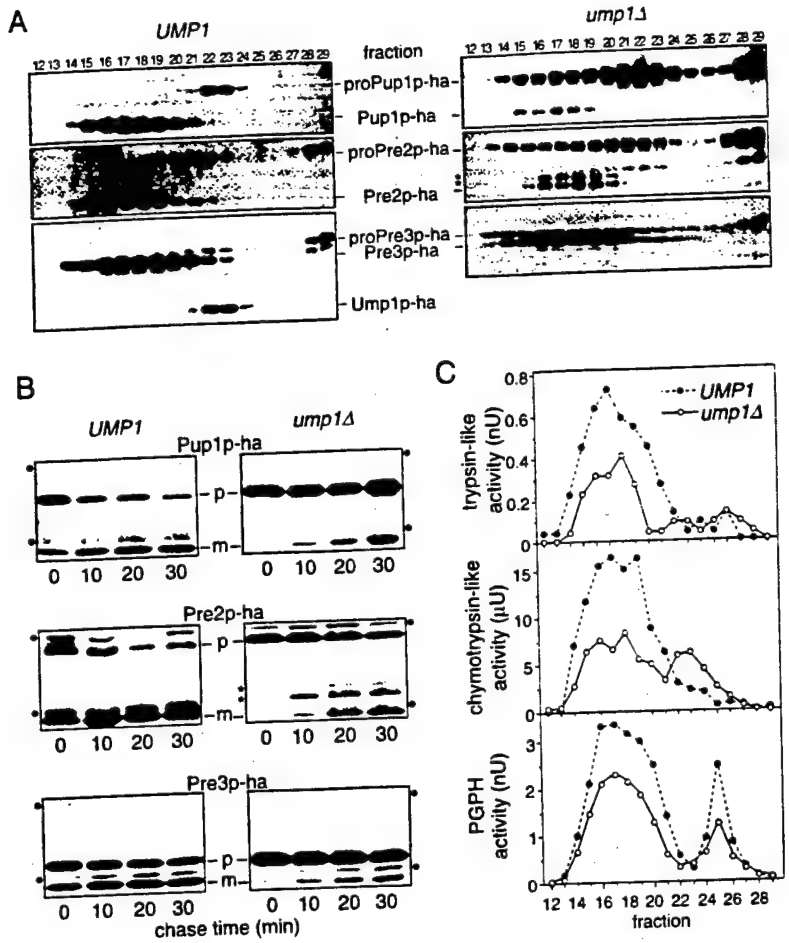



Figure 3. Ump1p Is Required for the Correct Processing of Proteasomal β Subunits

(A) SDS-PAGE and anti-ha immunoblot analyses of extracts fractionated on Superose-6. The congenic yeast strains used were JD131, JD132, JD133, JD134, JD135, and JD136 (Table 1). Western blots of the *ump1Δ* extracts were developed for a shorter time to compensate for an increased expression of proteasomal subunits in the mutant. Asterisks, processing intermediates or incompletely processed forms of Pre2p-ha that were observed in the *ump1Δ* mutant but not in wt cells.

(B) Pulse-chase analysis comparing the rate of β subunit processing in wt cells to that in *ump1Δ* cells (same strains as in A). Cells were pulse-labeled with $^{35}$S-Met/Cys for 5 min. Proteins were precipitated with anti-ha antibody. Fluorographic exposures were ~3 times longer for the *UMP1* samples. P, propeptide-containing precursor form; m, mature form; asterisks, as in (A); dots, nonspecific bands present in some samples.

(C) Comparison of proteolytic activities in extracts with equal amounts of total protein of *UMP1* (JD127) and *ump1Δ* cells (JD75) fractionated by gel filtration on Superose-6. The profile of the chymotrypsin-like activity in *UMP1* is the same as in Figure 2A.

containing the ~15S proteasome precursor from wt cells, and no processing intermediates of Pre2p could be detected either (Figure 3A). These findings indicate that Ump1p is required to prevent premature processing of at least proPre2p and that the presence of Ump1p is important for the coordination of proteasome assembly and subunit processing.

A defect of *ump1Δ* cells in the processing of β subunits was also observed by following the metabolic fate of these subunits in pulse-chase experiments. In wt cells, most of the propeptide-containing forms of β subunits (proPup1-ha, proPre2p-ha and proPre3p-ha) were converted into their mature counterparts during the 30-min chase (Figure 3B). By contrast, in the *ump1Δ* mutant, the bulk of the β subunit precursors remained unprocessed during this time, a finding consistent with the results of immunoblot analyses (Figure 3A). In the case of Pre2p-ha, pulse-chase analysis again revealed the species of intermediate size (putative processing intermediates) in extracts from *ump1Δ* cells. These species were absent from the equivalent samples derived from wt cells (Figure 3B).

### Ump1p Is Degraded upon Formation of the 20S Proteasome

The experiments above demonstrated that Ump1p is a component of proteasome precursors that is absent from the mature 20S proteasome. One possibility was that Ump1p leaves the precursor complex upon formation of the 20S proteasome from the two ~15S precursors and may then assist with another round of proteasome assembly, thus acting catalytically. It was also possible that formation of the 20S proteasome sterically traps Ump1p. Since the formation of the 20S form of the proteasome coincides with the appearance of its proteolytic activities (Frentzel et al., 1994; Chen and Hochstrasser, 1996), Ump1p might be degraded by the newly formed 20S proteasome. Pulse-chase analysis of Ump1p-ha in wt cells showed that Ump1p-ha is indeed degraded in vivo, the rate of its degradation being similar to the rate of disappearance of proPup1p-ha, which is converted to Pup1p-ha upon maturation of the proteasome (compare Figures 4A and 3B).

### Ump1p Is Stabilized in Proteasome Mutants and Can Be Detected Inside the 20S Particle

If Ump1p becomes a substrate of the newly formed 20S proteasome, the degradation of Ump1p should be inhibited by mutations that affect the proteasome's proteolytic activities. We used pulse-chase assays to follow the metabolic fate of Ump1p-ha in the *pre1-1* mutant, which is known to be deficient in the chymotryptic activity of the proteasome (Heinemeyer et al., 1993). Ump1p was partially stabilized in *pre1-1* cells, as indicated by the increased amount of $^{35}$S-labeled Ump1p in *pre1-1*

extracts versus wt extracts that contained the same total amount of TCA-precipitable $^{35}S$ (Figure 4A). However, we still observed a significant decrease of pulse-labeled Ump1p upon increasing chase times when the immunoprecipitation was carried out under nondenaturing conditions (Figure 4A). By contrast, when the extracts were treated with 0.4% SDS at 100°C before immunoprecipitation, virtually no decrease of the Ump1p signal during the chase was observed in *pre1-1* cells. These results were consistent with the possibility that the newly formed Ump1p-ha became inaccessible to the anti-ha antibody during the chase because it became trapped within the newly formed *pre1-1* proteasome. If so, Ump1p was expected to be present in fractions from the Superose-6 column that corresponded to the 20S and 26S proteasomes in the *pre1-1* mutant. Indeed, whereas in wt (*PRE1*) cells Ump1p-ha was detected by SDS-PAGE and immunoblotting only in fractions 22 and 23 (corresponding to the ~15S proteasome precursor complex), in the mutant (*pre1-1*) cells Ump1p-ha was detected in fractions 15–23, indicating that it was also present in mature proteasomes (Figure 4B). We assayed the accessibility of Ump1p-ha to anti-ha antibody by immunoprecipitation from fraction 19 (the 20S proteasome) of extracts from the *pre1-1* mutant and from fraction 22 (the ~15S proteasome precursor) of extract from *PRE1* cells. Ump1p-ha could be immunoprecipitated from either fraction 19 or 22 following pretreatment with 0.1% SDS, but not under nondenaturing conditions (Figure 4C), indicating that the ha tag was inaccessible to the anti-ha antibody in both the *pre1-1* proteasome and the precursor complex. However, when otherwise identical assays were carried out with a polyclonal anti-Ump1p antiserum, most of Ump1p could be immunoprecipitated under nondenaturing conditions from fraction 22 (the ~15S proteasome precursor) but was still not immunoprecipitated from fraction 19 (the 20S *pre1-1* proteasome). In contrast, when polyclonal anti-proteasome antibodies were used, Ump1p was precipitated quantitatively along with both complexes (Figure 4C).

These results indicated that a part of Ump1p other than its C terminus is accessible to anti-Ump1p antibodies in the ~15S precursor. However, Ump1p becomes entirely inaccessible (under nondenaturing conditions) upon formation of the 20S proteasome from two ~15S precursors. This interpretation was supported by examining the sensitivity of Ump1p-ha in different complexes to trypsin digestion (Figure 4D). Specifically, Ump1p-ha in the 20S proteasomes (fraction 19) from *pre1-1* cells was completely protected against trypsin. In contrast, Ump1p-ha in the ~15S proteasome precursors was detectably accessible to trypsin. The degradation of Ump1p was incomplete, resulting in a protected fragment of Ump1p-ha that lacked the ~5 kDa N-terminal region but retained the C-terminal ha tag (Figure 4D). Interestingly, in a similar experiment with a strain expressing pro-Pre2p-ha, we observed that, in the Ump1-containing ~15S proteasome precursor complex (fraction 22), pro-Pre2p-ha was shortened by trypsin treatment to yield a product whose electrophoretic mobility was indistinguishable from that of the natural mature Pre2p-ha (Figure 4D). In these experiments, the overall structure of the Ump1p proteasome precursor complex remained intact as judged by native gel analysis of trypsin-treated material (data not shown).

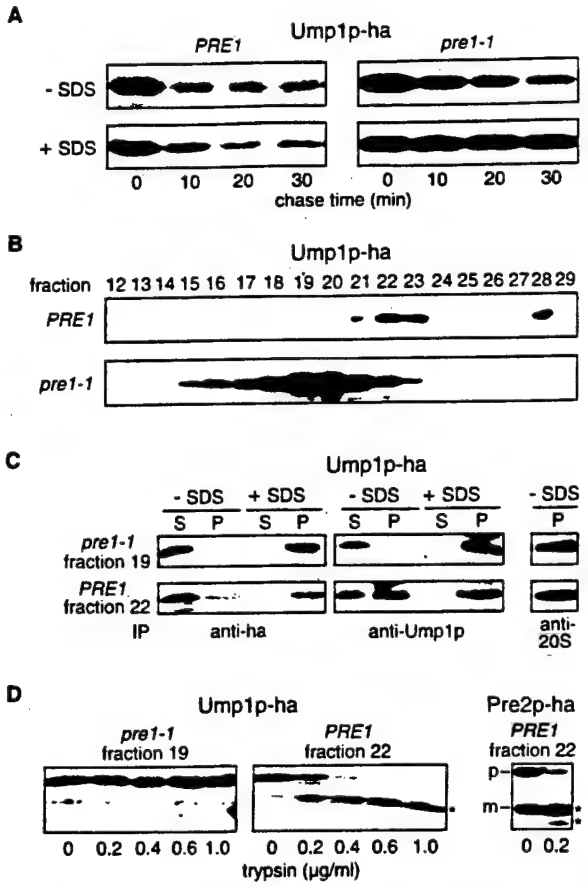



Figure 4. Ump1p Is Stabilized in *pre1-1* Mutants Cells that Are Defective in the Proteasome's Chymotrypsin-like Activity and Persists in 20S Proteasomes

(A) Pulse-chase analysis of Ump1p-ha in wt and *pre1-1* cells. The strains used were JD150 (*PRE1*) and JD151 (*pre1-1*) (Table 1), both expressing Ump1p-ha. The cells were pulse-labeled with $^{35}S$-Met/Cys for 5 min. Extracts were prepared from samples taken at different chase times. Samples were then split in halves. One-half (−SDS) was subjected to immunoprecipitation with anti-ha antibody according to the standard pulse-chase protocol. The other half (+SDS) was adjusted to 0.4% SDS and incubated at 100°C for 5 min, then diluted with extraction buffer to the final concentration of 0.1% SDS prior to immunoprecipitation.

(B–D) SDS-PAGE and anti-ha immunoblot analyses. (B) Detection of Ump1p-ha-containing complexes in extracts fractionated by Superose-6 gel filtration. Strains were the same as in (A). Note the accumulation of Ump1p-ha in fractions containing the 20S and 26S proteasome (fractions 15–21) in *pre1-1*. (C) Immunoblot analyses of immunoprecipitations were carried out with material from Superose-6 fraction 19 of extracts from strain JD151 and fraction 22 of extracts from strain JD150, using the indicated antibodies. Immunoprecipitations were performed without addition of SDS (−SDS) or after boiling in the presence of 0.1% SDS (+SDS). P, precipitates; S, supernatants. (D) Assaying trypsin sensitivity of the same material as in (C), and of the Superose-6 fraction 22, from strain JD138 expressing Pre2p-ha. P, proPre2p-ha; m, mature Pre2p-ha; asterisks mark trypsin digestion products.

## In the *ump1Δ* Mutant, the Propeptide of Pre2p Is Not Required for Incorporation of Pre2p into the Proteasome

Chen and Hochstrasser (1996) have elegantly demonstrated that the propeptide of Pre2p, if separated from the mature Pre2p, could function in *trans* and thereby

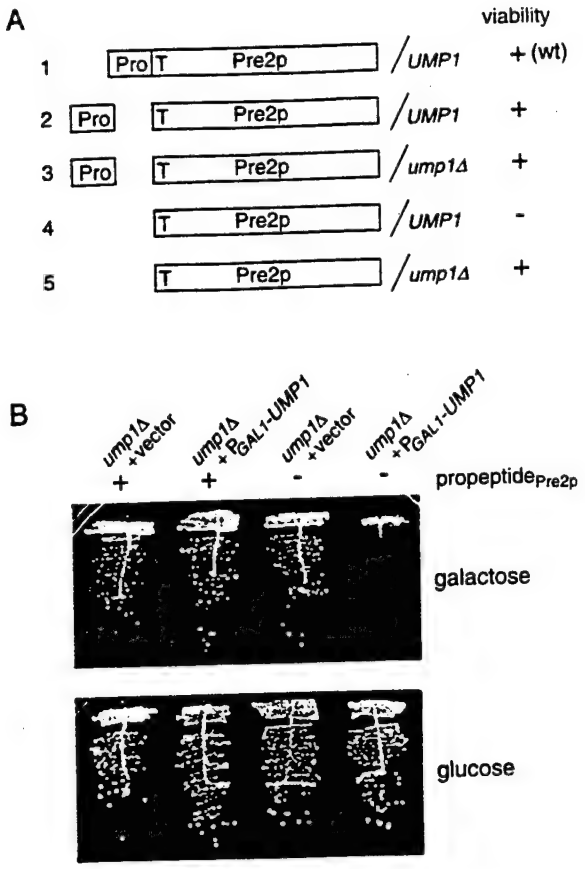



Figure 5. The Pre2p Propeptide Is Not Essential in the *ump*Δ Mutant

(A) Representation of different genotypes of a set of congenic strains and their effects on cell viability. Strain 1 (JD50-10B) is the wt control. Strain 2 (MHY952) expresses Pre2p/Doa3p-ΔLS with N-terminal Thr (T) as a fusion to ubiquitin (not shown), and the Pre2p/Doa3p propeptide (LS) from two separate plasmids (Chen and Hochstrasser, 1996). Strain 3 (JD160), same as strain 2 but *ump1*Δ. Strain 4 would be derived from strain 2 through the loss of the plasmid encoding Pre2p/Doa3p-LS that operates in *trans*. We were unable to produce isolates that lost this plasmid, confirming the earlier demonstration that the propeptide is essential for cell viability (Chen and Hochstrasser, 1996). Strain 5 (JD163) is derived from strain 3 through the loss of the propeptide-encoding plasmid. Loss of the plasmid did not affect cell viability in the *ump1*Δ background.

(B) Induction of *UMP1* expression in the *ump1*Δ background inhibits the growth of a strain that lacks the Pre2p propeptide. Strains 3 (plus) and 5 (minus) shown in (A) were transformed with pJD429 that expressed *UMP1* from the galactose-inducible, glucose-repressible $P_{GAL1}$ promoter, or with an empty vector as a control. Transformants selected on glucose media were pregrown on selective media containing raffinose as a carbon source, then streaked onto selective media with 2% galactose or 2% glucose, and grown for 3 days at 30°C.

still allow the incorporation of Pre2p into the 20S proteasome. Under these conditions, expression of the propeptide is essential for cell viability, suggesting that the propeptide, in *cis* or at least in *trans*, is required for the assembly of an active proteasome (Chen and Hochstrasser, 1996).

We confirmed this result using their strain MHY952, in which mature Pre2p and the propeptide are expressed on separate plasmids (Figure 5A). Specifically, under nonselective growth conditions this strain did not yield viable cells that had lost the plasmid expressing the propeptide. Surprisingly, however, when we constructed and examined a congenic *ump1*Δ derivative of this strain, we noticed that it lost the propeptide-expressing plasmid at a frequency suggesting that this plasmid did not provide a significant growth advantage to cells (data not shown). Indeed, growth rates of cells with the plasmid were indistinguishable from those lacking it (Figure 5B). This result demonstrated that *ump1*Δ is a suppressor of a deletion of the propeptide of Pre2p. If so, expression of *UMP1* in the (*ump1*Δ *PRE2*-Δ*pro*) background should be incompatible with cell viability. This prediction was confirmed when we examined growth properties of the (*ump1*Δ *PRE2*-Δ*pro*) strain transformed with a plasmid expressing *UMP1* from the galactose-inducible $P_{GAL1}$ promoter. On glucose-containing media, the growth rate of this transformant was indistinguishable from that of an otherwise identical transformant carrying the propeptide-expressing plasmid. In contrast, on galactose-containing media, the (*ump1*Δ *PRE2*-Δ*pro*) mutant containing the $P_{GAL1}$-*UMP1* plasmid was unable to grow, whereas the control strains grew (Figure 5B). This result indicated that the propeptide of Pre2p is essential for the Ump1p-assisted maturation of the proteasome but is not essential for (partially) defective maturation of the proteasome that takes place in the absence of Ump1p.

## Discussion

### Ump1p, a Novel Maturation Factor of the 20S Proteasome

We describe the discovery of a proteasome maturation factor, termed Ump1p, whose unique properties include a short in vivo half-life that is due to its degradation within the newly formed proteasome. Ump1p is required for coordination of the proteasome's physical assembly and enzymatic activation. In addition, the normally essential propeptide of the Pre2p β subunit was found to become nonessential in the absence of Ump1p. We report the following specific results.

#### ump1Δ *Mutants Are Defective in Ub-Mediated Proteolysis*

They are sensitive to a variety of stresses and accumulate Ub-protein conjugates (Figure 1 and data not shown). All of the observed phenotypes of *ump1*Δ cells are consistent with the conclusion that *ump1* mutants are impaired in proteasome biogenesis and, consequently, in the degradation of ubiquitylated proteins.

#### *The* UMP1 *Gene*

It encodes a polypeptide with a calculated molecular mass of 16.8 kDa. No close sequence homologs of Ump1p were detected in the current databases. However, the presence of a small protein similar in size to Ump1p in preparations of the half-proteasome precursors in mammals (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidtke et al., 1997) suggests the presence of a functional homolog of Ump1p in the mammalian proteasome maturation pathway.

#### *Ump1p Is a Component of Proteasome Precursors that Contain Unprocessed* β *Subunits*

The Ump1p-containing complex has a molecular mass of 300–400 kDa, sediments at ~15S, migrates significantly faster in native gels than the 20S proteasome,

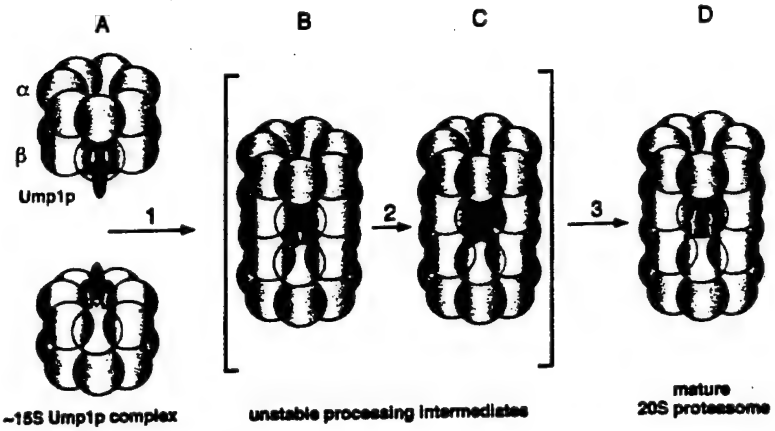



Figure 6. A Model of the Ump1p Function in Proteasome Maturation

Shown is a schematic view of the 20S proteasome and its precursor forms, with the α and β subunits as blue and green balls, respectively. Three of the β subunits are drawn with extensions that represent propeptides. The β subunits in the front are drawn transparent in order to allow a view into the interior chamber of the proteasome. Structure A is a proteasome precursor complex (half proteasome), characterized by the presence of Ump1p and unprocessed β subunits. In step 1, two of these precursors join to build structure B, a step that leads to conformational or positional shifts of Ump1p and propeptides. Conformational changes of the propeptides trigger their autocatalytic processing (step 2), and activation of the proteasome's proteolytic activities. This leads to the degradation of the chamber-entrapped Ump1p by the newly formed proteasome (step 3). Only structures A and D are long-lived enough to be detected in wt cells. The findings with the *pre1-1* mutant, in which the degradation of Ump1p is inhibited, suggest the existence of the short-lived intermediates B and C. See text for details.

and has a subunit composition that is highly similar to that of the 20S proteasome, as determined by SDS-PAGE and immunoblot analyses with anti-20S proteasome antibodies (Figure 2 and data not shown). Taken together, these results indicate that the Ump1p is contained within a half-proteasome precursor complex, whose counterpart has been described in the maturation pathway of mammalian proteasomes (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidtke et al., 1997).

### *Maturation of Pup1p, Pre2p, and Pre3p Is Strongly Impaired in the* ump1Δ *Mutant*

Considerably increased amounts of the unprocessed (propeptide-containing) forms of these β subunits were detected in the 20S and 26S fractions from *ump1*Δ cells (Figure 3). In addition, the proteasomes assembled in *ump1*Δ cells had reduced activity. Thus, Ump1p is required for the correct and efficient maturation of the proteasomes.

### *Partially Processed Forms of Pro-Pre2p Are Detected in the Proteasome Precursor Complex and the Proteasome of* ump1Δ *Cells but Not in WT Cells*

These incompletely and prematurely processed forms of Pre2p (Figures 3A and 3B) appear to underlie a chymotrypsin-like activity that was associated with the precursor complex in *ump1*Δ cells but was absent from the same complex of wt cells (Figure 3C). This was suggested by the observation that this activity was not inhibited by treatment with lactacystin and was absent from cells lacking the Pre2p propeptide (P. C. R. and R. J. D., unpublished data). These findings indicated that Ump1p has a dual role in proteasome maturation. Specifically, Ump1p prevents premature processing of proPre2p in the precursor complex and is also required for the correct maturation of active sites upon assembly of the proteasome (see below).

### *Ump1p Is Degraded by the Newly Formed Proteasome*

The kinetics of the rapid degradation of Ump1p is similar to the kinetics of processing of Pup1p (Figures 3B and 4A), suggesting that Ump1p is destroyed by the proteasome upon its formation from the two half-proteasome precursors and the accompanying maturation of active sites (Chen and Hochstrasser, 1996). This conclusion was strongly supported by the observation that Ump1p is significantly stabilized in the *pre1-1* mutant, which is deficient in the proteasome's chymotrypsin-like activity (Figure 4A). Specifically, in contrast to the pattern in wt cells, Ump1p was detectable in the 20S and 26S forms of the *pre1-1* proteasome. Ump1p in these complexes was shielded both from detection by antibodies and from digestion by trypsin (Figures 4B–4D). These and related data suggested a model in which Ump1p is encased within the 20S proteasome upon its assembly from the two Ump1p-containing half-proteasome precursor complexes (Figure 6). Formation of the 20S structure triggers active site maturation, resulting in the degradation of Ump1p.

### *The Propeptide of Pre2p Is Not Essential in* ump1Δ *Cells*

Chen and Hochstrasser (1996) demonstrated that the propeptide of Pre2p is essential for cell viability and that this propeptide can operate in *trans*. They concluded that the propeptide, in addition to rendering proteasome precursors proteolytically inactive, serves a chaperone-like function required for the correct incorporation of Pre2p into the proteasome. One striking result of the present work is that the propeptide of Pre2p becomes dispensable in the *ump1*Δ mutant (Figure 5). In fact, the presence of the propeptide, which is essential for viability of wt cells, does not provide a growth advantage to *ump1*Δ cells that express it.

## On the Functions of Ump1p and β Subunit Propeptides in Proteasome Maturation

Our results suggested a model illustrated in Figure 6. This model is based in part on the idea described by Chen and Hochstrasser (1996)—that the active sites of the proteasome are formed upon its assembly from two half-proteasome precursors, through a juxtaposition of subunits at the interface of the dyad-related halves. The advantage of this mechanism is that autocatalytic maturation of the active sites (through processing of the relevant β subunits) is coupled to the assembly of the proteasome, thus avoiding premature processing of the propeptides of β subunits. This view was supported by

Table 1. Yeast Strains

| Strain | Relevant Genotype | Source/Comment |
|---|---|---|
| YPH500 | *MATα ade2-101 his3-Δ200 leu2-Δ1 lys2-801 trp1-Δ63 ura3-52* | Sikorski and Hieter, 1989 |
| JDC9-11 | *MATα ump1-1* | Derivative of YPH500 |
| JD47-13C | *MATa his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52* | Dohmen et al., 1995 |
| JD59 | *MATa ump1-Δ1::HIS3* | Derivative of JD47-13C |
| JD53 | *MATα* | Derivative of JD47-13C |
| JD81-1A | *MATα ump1-Δ1::HIS3* | Derivative of JD53 |
| JD75 | *MATa PRE1-ha::Yiplac211 ump1-Δ1::HIS3* | Derivative of JD47-13C |
| JD126 | *MATa UMP1-ha::Yiplac128 PRE1-Flag-6His::Yiplac211* | Derivative of JD47-13C |
| JD127 | *MATa UMP1-ha::Yiplac128 PRE1-ha::Yiplac211* | Derivative of JD71 |
| JD129 | *MATa UMP1-ha::Yiplac128* | Derivative of JD127 |
| JD131 | *MATa UMP1-ha::Yiplac128 PUP1-ha::Yiplac211* | Derivative of JD129 |
| JD132 | *MATa ump1-Δ1::HIS3 PUP1-ha::Yiplac211* | Derivative of JD59 |
| JD133 | *MATa UMP1-ha::Yiplac128 PRE2-ha::Yiplac211* | Derivative of JD129 |
| JD134 | *MATa ump1-Δ1::HIS3 PRE2-ha::Yiplac211* | Derivative of JD59 |
| JD135 | *MATa UMP1-ha::Yiplac128 PRE3-ha::Yiplac211* | Derivative of JD129 |
| JD136 | *MATa ump1-Δ1::HIS3 PRE3-ha::Yiplac211* | Derivative of JD59 |
| JD138 | *MATa PRE2-ha::Yiplac211* | Derivative of JD47-13C |
| JD139 | *MATa PUP1-ha::Yiplac211* | Derivative of JD47-13C |
| BBY45 | *MATa his3-Δ200 leu2-3,112 lys2-801 trp1-1 ura3-52* | Bartel et al., 1990 |
| JD61 | *MATa/MATα ump1-Δ1::HIS3/ump1-Δ2::LEU2* | Congenic with BBY45 |
| JD50-10B | *MATα leu2-3,112 his3-Δ200 trp1-1 ura3-52* | Congenic with BBY45 |
| MHY952 | *MATα leu2-3,112 his3-Δ200 trp1-1 ura3-52 pre2(=doa3)Δ1::HIS3* (YCpUbDOA3ΔLS-His) ($YEpDOA3_{LS}$) | Chen and Hochstrasser, 1996, congenic with BBY45 |
| JD160 | *MATα ump1-Δ3 pre2(=doa3)Δ1::HIS3* (YCpUbDOA3ΔLS-His) ($YEpDOA3_{LS}$) | Derivative of MHY952 |
| JD163 | *MATα ump1-Δ3 pre2(=doa3)-Δ1::HIS3* (YCpUbDOA3ΔLS-His) | Derivative of JD160, cured of $YEpDOA3_{LS}$ |
| WCG4a | *MATα ura3 his3-11 leu2-3,112* | Heinemeyer et al., 1993 |
| YHI29/1 | *MATα pre1-1 ura3 his3-11 leu2-3,112* | Heinemeyer et al., 1993 |
| JD150 | *MATα UMP1-ha::Yiplac128* | Derivative of WCG4a |
| JD151 | *MATα pre1-1 UMP1-ha::Yiplac128* | Derivative of YH129/1 |

biochemical analyses of the proteasome maturation in mammalian cells (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidtke et al., 1997).

The results of the present work further support this model. In addition, we discovered that a novel factor, Ump1p, is required for autocatalytic active site maturation and proteasome assembly to occur in a coordinated fashion. A model that accounts for our findings and is consistent with the available evidence is illustrated in Figure 6. In this model, Ump1p has a chaperone-like function required for the efficient processing of the propeptides of β subunits upon the assembly of 20S proteasomes. Specifically, Ump1p interacts with these β subunits, most likely with their propeptides (structure A in Figure 6). Upon formation of the 20S proteasome (Figure 6, step 1), a conformational change of Ump1p, or a change of its position within the complex, would induce a conformational change of the propeptide that results in autocatalytic processing (Figure 6, step 2).

Why is the propeptide of Pre2p essential for proteasome assembly and activation *only* in the presence of Ump1p? We propose that the propeptide of Pre2p is required to induce an alteration of conformation or position of Ump1p upon the assembly of the 20S particle. In this model, the absence of the propeptide would leave Ump1p in a position that is incompatible with the formation of active proteasome. It is possible (and remains to be verified) that this chaperone-like function of a propeptide is unique to Pre2p, as it contains a much longer propeptide (75 residues) than the other β subunits.

In the model of Figure 6, Ump1p is a metabolically unstable chaperone that is required for the proper (and properly timed) processing of β subunits upon the assembly of the 20S proteasome and that is destroyed within the newly activated proteasome. The unusual mechanics of Ump1p, its noncatalytic mode of action, and its degradation by the protease it helps to activate characterize Ump1p as a novel type of molecular chaperone.

## Experimental Procedures

### Yeast Media

Yeast rich (YPD) and synthetic (S) minimal media with 2% dextrose (SD) or 2% galactose (SG) were prepared as described (Dohmen et al., 1995).

### Isolation of the *UMP1* Gene

*UMP1* was identified using a selection-based screen for mutants in the N-end rule pathway (reviewed by Varshavsky, 1996). *S. cerevisiae* strain YPH500 (Table 1) was transformed with two plasmids: pJD205, which expressed a Ura3p-based N-end rule reporter substrate (Arg-Tpi1p-Ura3p), and pRL2, which expressed Arg–β-gal, another N-end rule reporter substrate that can be monitored using X-Gal plate assays (Baker and Varshavsky, 1995). Previous work (Dohmen et al., 1994) has shown that rapid degradation of a Ura3-based N-end rule substrate renders the cells $Ura^-$, whereas mutants in the N-end rule pathway that express the same Ura3p-based reporter are $Ura^+$. In this screen, $Ura^-$ isolates were selected on SD media lacking uracil and were then tested on X-Gal plates to verify that the same isolates were also defective in the degradation of the Arg–β-gal N-end rule substrate. One of the mutants thus identified (*ump1-1*) was cured of pJD205 and transformed with a genomic yeast library (Rose et al., 1987). Six transformants yielded plasmids with four overlapping inserts that could restore the ability of cells to degrade Arg–β-gal. A 1981 bp BamHI/Sau3A fragment common

to all of the inserts was sequenced and found to contain one complete ORF and two flanking incomplete ORFs. Further mapping, using subcloning and PCR, confirmed that the complete ORF of 445 bp was responsible for the complementation. The *UMP1* sequence (EMBL database accession number: AJ002557) is identical to ORF YBR173C, an ORF subsequently identified by the yeast genome project.

### Construction of Yeast Strains and Plasmids

Table 1 lists the strains used in this study. To construct *ump1*Δ alleles, the 1981 bp fragment described above was subcloned into M13mp19. Using single-stranded DNA of the resulting phage, a synthetic oligonucleotide and T4 DNA polymerase, the *UMP1* ORF was precisely deleted and replaced by a BglII restriction site. The resulting fragment was subcloned into pUC19, and the BglII site was used to insert fragments containing the *HIS3*, *LEU2*, or the *URA3* gene, the latter one being flanked by two direct repeats of a segment of the *E. coli hisG* gene (Alani et al., 1987). The resulting deletion alleles (*ump1-Δ1::HIS3*, *ump1-Δ2::LEU2*, and *ump1-Δ3::URA3*) were isolated as BamHI fragments, introduced into *S. cerevisiae*, and used to delete the *UMP1* gene (Rothstein, 1991). *ump1*Δ strains carrying an unmarked *ump1-Δ3* allele (resulting from recombination between the *hisG* repeats) were selected on plates containing 5-fluoroorotic acid (Alani et al., 1987). Construction of chromosomal ORFs that expressed C terminally tagged versions of Ump1p or proteasome subunits (Pre1p, Pre2p, Pre3p, and Pup1p) instead of their wt counterparts was performed as follows. Using primers that contained flanking EcoRI and KpnI sites, 3′ portions of the respective genes were amplified by PCR. These sites were then used to insert the amplified fragments into integrative plasmids based on YIplac128 (*LEU2* marked) or YIplac211 (*URA3* marked) (Gietz and Sugino, 1988) that contained sequences encoding epitope tags followed by the terminator sequence of the *CYC1* gene ($T_{CYC1}$). Each of the resulting plasmids was linearized within the coding sequence for targeted integration into the *S. cerevisiae* genome, yielding strains with one copy of the respective gene (fused in-frame to the tag-coding sequence) expressed from its natural promoter and $T_{CYC1}$, in addition to a 3′ portion of the same gene without promoter. The epitope tags used were double ha ("ha") and Flag-$His_6$ ("FH") (Dohmen et al., 1995). Plasmid pJD429 (*CEN/URA3*) expressing a *UMP1* cDNA from *PGAL1* was isolated from a cDNA library (Liu et al., 1992) through complementation of the *ump1*Δ mutation.

### Pulse-Chase Analyses

R–β-gal (Ub-R–β-gal) and Ub-P–β-gal (Bachmair et al., 1986) were expressed in the *MAT*a strains JD47-13C or JD59. MATα$2_{1-66}$–β-gal was expressed from the plasmid YCp50-α$2_{deg1}$–β-gal (Hochstrasser and Varshavsky, 1990) in the *MAT*α strains JD53 and JD81-1A (Table 1). Pulse labeling for 5 min with Redivue Promix [$^{35}$S] (Amersham) followed by a chase and immunoprecipitation with monoclonal anti–β-gal (Promega) or anti-ha (16B12, Babco) antibodies were carried out as described by Dohmen et al. (1991). $^{35}$S proteins fractionated by SDS-PAGE were detected by fluorography.

### Fractionation of Whole-Cell Extracts by Gel Filtration

*S. cerevisiae* were grown at 30°C in YPD or in SD media to $OD_{600}$ of 1.4 ± 0.1, harvested at 3000g, washed with cold water, frozen in liquid nitrogen, and stored at −80°C. Cell paste was ground to powder in a mortar in the presence of liquid nitrogen. The extraction buffer was 50 mM Tris-HCl (pH 7.5), 2 mM ATP, 5 mM $MgCl_2$, 1 mM DTT, 15% glycerol, used at 2 ml per gram of pelleted yeast cells. After centrifugation at 31,000g for 10 min at 2°C, the supernatant was subjected to a second centrifugation at 60,000g for 30 min, yielding an extract with the protein concentration of ~5 mg/ml. The protein concentration of extracts was equalized in parallel experiments, using the extraction buffer. Using the FPLC system (Pharmacia), 200 μl samples of an extract were chromatographed on a Superose-6 column equilibrated with extraction buffer. The flow rate was 0.3 ml/min and fractions of 0.6 ml were collected. The Superose-6 column was calibrated using the following standards: thyroglobulin (669 kDa), ferritin (443 kDa), and bovine serum albumin (66 kDa). Dextran blue was used to monitor the void volume.

### Assays for Proteolytic Activities with Fluorogenic Peptide Substrates

To determine the chymotrypsin-like activity, 20 μl of the protein fraction and 20 μl of 0.5 mM succinyl-Leu-Leu-Val-Tyr-7-amido-4-methylcoumarin in 50 mM Tris-HCl (pH 7.8), 2 mM ATP, 5 mM $MgCl_2$, and 1 mM DTT were mixed and incubated for 1 hr at 37°C. The reaction was stopped by addition of 960 μl of cold ethanol, and the fluorescence was measured at 440 nm, using the excitation wavelength of 380 nm. The trypsin-like activity and the peptidylglutamyl peptide-hydrolyzing activity were determined with, respectively, Cbz-Ala-Ala-Arg-4MeO-β-naphthylamide and Cbz-Leu-Leu-Glu-β-naphthylamide as fluorogenic peptide substrates (Fischer et al., 1994), except that the volume of the protein fraction was 75 μl, the reactions were stopped after 270 min at 37°C, and the fluorescence emission was measured at 366 nm, using an excitation wavelength of 420 nm. One unit (U) of a proteolytic activity is defined as 1 μmol of the fluorophore produced per min under these conditions.

### Immunoprecipitation, Electrophoresis, and Immunoblotting

Immunoprecipitations were carried out using either monoclonal anti-tag antibodies (see below), or polyclonal rabbit antisera raised against the yeast 20S proteasome (a gift from K. Tanaka), or against Ump1p-$His_6$ expressed in *E. coli*. For immunoblotting, the protein samples were boiled for 3 min in the presence of 2% SDS and 0.1 M 2-mercaptoethanol, then subjected to 12% SDS-PAGE, and thereafter transferred onto a PVDF membrane (Millipore) in a wet blot system (BioRad). The blots were incubated with either rabbit anti-ubiquitin (Ramos et al., 1995), or anti-Cim3p (Ghislain et al., 1993) polyclonal antibodies, or 16B12 anti-ha (BabCo), or M2 anti-Flag (Kodak) monoclonal antibodies, and were processed as described (Ramos et al., 1995), except that the initially blotted proteins were visualized using horseradish peroxidase-conjugated goat anti-mouse or anti-rabbit IgGs, the chemiluminescence blotting substrate detection system from Boehringer Mannheim, and X-ray films. Nondenaturing 4.5% acrylamide gel electrophoresis was performed as described by Hough et al. (1987), and the gels were incubated for 15 min in transfer buffer containing 0.1% SDS before electrotransfer.

### Affinity Purification of Proteasomes and Related Complexes

The 20S proteasome and its precursors were purified from strain JD126 that expressed Pre1p-FH and Ump1p-ha. Crude extracts were prepared as described above, followed by an exchange of buffer to 50 mM Na-phosphate (pH 8.0), 0.3 M NaCl, using PD-10 columns (Pharmacia). The extract was then incubated with Ni-NTA agarose (Qiagen) in batch for 2 hr at 4°C, followed by washings and elution according to the manufacturer's protocol, except that the (pH 6.0) washing buffer contained 20 mM imidazole. $His_6$-tagged proteins were eluted with a step gradient of 100-500 mM imidazole. Active fractions containing a mixture of the mature proteasome and its precursors were pooled, diluted 2-fold with 50 mM Tris-HCl (pH 7.5), and incubated for 2 hr at 4°C in batch with 0.5 ml anti-Flag M2 antibody affinity resin (IBI/Eastman Kodak) that had been equilibrated in TN buffer (0.15 M NaCl, 50 mM Tris-HCl, [pH 7.5]). After the loading, the resin was washed with TN buffer. Flag-tagged proteins were specifically eluted with the Flag epitope peptide (IBI/Eastman Kodak).

### Acknowledgments

We thank Rohan Baker, Anthony Bretscher, Ricardo Ferreira, Wolfgang Heinemeyer, Wolfgang Hilt, Mark Hochstrasser, Carl Mann, Nancy Kleckner, Keiji Tanaka, and Dieter Wolf for the gifts of plasmids, yeast strains, and antisera; Elisabeth Andrews for assistance in the cloning of *UMP1*, Reiner Stappen for assistance in producing anti-Ump1p antibodies, Robert Kramer (Limberg Druck) for printing the figures, and Nils Johnsson and Ralf Kölling for helpful suggestions. P. C. R., J. H., and R. J. D. are grateful to Cornelis P. Hollenberg for providing lab space and for his support, and to Isabel Fuchs for technical assistance. P. C. R. was supported by a postdoctoral fellowship from Fundação para a Ciência e Tecnologia, Programa Praxis XXI. This work was supported by a grant to R. J. D. from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (0316711), by start-up funding from the Ministerium für

Wissenschaft und Bildung des Landes Nordrhein-Westfalen, and by a grant to A. V. from the National Institutes of Health (GM31530).

Received December 10, 1997; revised January 16, 1998.

## References

Alani, E., Cao, L., and Kleckner, N. (1987). A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains. Genetics *116*, 541–545.

Arendt, C.S., and Hochstrasser, M. (1997). Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. Proc. Natl. Acad. Sci. USA *94*, 7156–7161.

Bachmair, A., Finley, D., and Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. Science *234*, 179–186.

Baker, R.T., and Varshavsky, A. (1995). Yeast N-terminal amidase—a new enzyme and component of the N-end rule pathway. J. Biol. Chem. *270*, 12065–12074.

Bartel, B., Wünning, I., and Varshavsky, A. (1990). The recognition component of the N-end rule pathway. EMBO J. *9*, 3179–3189.

Chen, P., and Hochstrasser, M. (1996). Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. Cell *86*, 961–972.

Coux, O., Tanaka, K., and Goldberg, A.L. (1996). Structure and functions of the 20S and 26S proteasomes. Annu. Rev. Biochem. *65*, 801–847.

Deveraux, Q., Ustrell, V., Pickart, C., and Rechsteiner, M. (1994). A 26S protease subunit that binds ubiquitin conjugates. J. Biol. Chem. *269*, 7059–7061.

Dohmen, R.J., Madura, K., Bartel, B., and Varshavsky, A. (1991). The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme. Proc. Natl. Acad. Sci. USA *88*, 7351–7355.

Dohmen, R.J., Wu, P., and Varshavsky, A. (1994). Heat-inducible degron: a method for constructing temperature-sensitive mutants. Science *263*, 1273–1276.

Dohmen, R.J., Stappen, R., McGrath, J.P., Forrova, H., Kolarov, J., Goffeau, A., and Varshavsky, A. (1995). An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. J. Biol. Chem. *270*, 18099–18109.

Fischer, M., Hilt, W., Richter-Ruoff, B., Gonen, H., Ciechanover, A., and Wolf, D.H. (1994). The 26S proteasome of the yeast *Saccharomyces cerevisiae*. FEBS Lett. *355*, 69–75.

Frentzel, S., Pesold-Hurt, B., Seelig, A., and Kloetzel, P.-M. (1994). 20S proteasomes are assembled via distinct precursor complexes. J. Mol. Biol. *236*, 975–981.

Ghislain, M., Udvardy, A., and Mann, C. (1993). *S. cerevisiae* 26S protease mutants arrest cell division in G2/metaphase. Nature *366*, 358–362.

Gietz, R.D., and Sugino, A. (1988). New yeast-*Escherichia coli* shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. Gene *74*, 527–534.

Groll, M., Ditzel, L., Loewe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. Nature *386*, 463–471.

Heinemeyer, W., Gruhler, A., Mohrle, V., Mahe, Y., and Wolf, D.H. (1993). *PRE2*, highly homologous to the human major histocompatibility complex-linked *RING10* gene, codes for a yeast proteasome subunit necessary for chymotryptic activity and degradation of ubiquitinated proteins. J. Biol. Chem. *268*, 5115–5120.

Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U., and Wolf, D.H. (1997). The active sites of the eukaryotic 20S proteasome and their involvement in subunit precursor processing. J. Biol. Chem. *272*, 25200–25209.

Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. Annu. Rev. Genet. *30*, 405–439.

Hochstrasser, M., and Varshavsky, A. (1990). In vivo degradation of a transcriptional regulator: the yeast α2 repressor. Cell *61*, 697–708.

Hochstrasser, M., Papa, F.R., Chen, P., Swaminathan, S., Johnson, P., Stillman, L., Amerik, A.Y., and Li, S.J. (1995). The doa pathway—studies on the functions and mechanisms of ubiquitin-dependent protein degradation in the yeast *Saccharomyces cerevisiae*. Cold Spring Harb. Symp. Quant. Biol. *60*, 503–513.

Hough, R., Ratt, G., and Rechsteiner, M. (1987). Purification of two high molecular weight proteases from rabbit reticulocyte lysate. J. Biol. Chem. *262*, 8303–8313.

Jentsch, S., and Schlenker, S. (1995). Selective protein degradation—a journey's end within the proteasome. Cell *82*, 881–884.

Johnson, E.S., Ma, P.C., Ota, I.M., and Varshavsky, A. (1995). A proteolytic pathway that recognizes ubiquitin as a degradation signal. J. Biol. Chem. *270*, 17442–17456.

Jungmann, J., Reins, H.A., Schobert, C., and Jentsch, S. (1993). Resistance to cadmium mediated by ubiquitin-dependent proteolysis. Nature *361*, 369–371.

Liu, H., Krizek, J., and Bretscher, A. (1992). Construction of a GAL1-regulated yeast cDNA expression library and its application to the identification of genes whose overexpression causes lethality in yeast. Genetics *132*, 665–673.

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. Science *268*, 533–539.

Lupas, A., Flanagan, J.M., Tamura, T., and Baumeister, W. (1997). Self-compartmentalizing proteases. Trends Biochem. Sci. *22*, 399–404.

Nandi, D., Woodward, E., Ginsburg, D.B., and Monaco, J. (1997). Intermediates in the formation of mouse 20S proteasomes: implications for the assembly of precursor β subunits. EMBO J. *16*, 5363–5375.

Peters, J.M. (1994). Proteasomes: protein degradation machines of the cell. Trends Biochem. Sci. *19*, 377–382.

Ramos, P., Cordeiro, A., Ferreira, R., Ricardo, C., and Teixeira, A. (1995). The presence of ubiquitin-protein conjugates in plant chloroplast lysates is due to cytosolic contamination. Austr. J. Plant Physiol. *22*, 893–901.

Rose, M.D., Novick, P., Thomas, J.H., Botstein, D., and Fink, G.R. (1987). A *Saccharomyces cerevisiae* genomic plasmid bank based on a centromere-containing shuttle vector. Gene *60*, 237–243.

Rothstein, R. (1991). Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol. *194*, 281–301.

Schmidtke, G., Schmidt, M., and Kloetzel, P.-M. (1997). Maturation of mammalian 20S proteasome: purification and characterization of 13S and 16S proteasome precursor complexes. J. Mol. Biol. *268*, 95–106.

Seemüller, E., Lupas, A., Stock, D., Löwe, J., Huber, R., and Baumeister, W. (1995). Proteasome from *Thermoplasma acidophilum*: a threonine protease. Science *268*, 579–582.

Seemüller, E., Lupas, A., and Baumeister, W. (1996). Autocatalytic processing of the 20S proteasome. Nature *382*, 468–470.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics *122*, 19–27.

Varshavsky, A. (1996). The N-end rule—functions, mysteries, uses. Proc. Nat. Acad. Sci. USA *93*, 12142–12149.

Varshavsky, A. (1997). The ubiquitin system. Trends Biochem. Sci. *22*, 383–387.

Yang, Y., Früh, K., Ahn, K., and Peterson, P.A. (1995). In vivo assembly of the proteasomal complexes, implications for antigen processing. J. Biol. Chem. *270*, 27687–27694.

*Proc. Natl. Acad. Sci. USA*
Vol. 95, pp. 7898–7903, July 1998
Biochemistry

# The mouse and human genes encoding the recognition component of the N-end rule pathway

**(ubiquitin/proteolysis/E3/N-recognin/*Ubr1*)**

YONG TAE KWON*, YUVAL REISS†, VICTOR A. FRIED‡, AVRAM HERSHKO§, JEONG KYO YOON*, DAVID K. GONDA¶, PITCHAI SANGAN¶, NEAL G. COPELAND‖, NANCY A. JENKINS‖, AND ALEXANDER VARSHAVSKY*,**

*Division of Biology, California Institute of Technology, Pasadena, CA 91125; †Department of Biochemistry, Tel Aviv University, Tel Aviv 69978, Israel; ‡Department of Cell Biology and Anatomy, New York Medical College, Valhalla, NY 10595; §Unit of Biochemistry, Faculty of Medicine, Technion, Haifa 31096, Israel; ¶Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06520-8024; and Mammalian Genetics Laboratory, Advanced BioScience Laboratories-Basic Research Program, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, MD 21702

*Contributed by Alexander Varshavsky, May 5, 1998*

**ABSTRACT** **The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. The N-end rule pathway is one proteolytic pathway of the ubiquitin system. The recognition component of this pathway, called N-recognin or E3, binds to a destabilizing N-terminal residue of a substrate protein and participates in the formation of a substrate-linked multiubiquitin chain. We report the cloning of the mouse and human *Ubr1* cDNAs and genes that encode a mammalian N-recognin called E3α. Mouse UBR1p (E3α) is a 1,757-residue (200-kDa) protein that contains regions of sequence similarity to the 225-kDa Ubr1p of the yeast *Saccharomyces cerevisiae*. Mouse and human UBR1p have apparent homologs in other eukaryotes as well, thus defining a distinct family of proteins, the UBR family. The residues essential for substrate recognition by the yeast Ubr1p are conserved in the mouse UBR1p. The regions of similarity among the UBR family members include a putative zinc finger and RING-H2 finger, another zinc-binding domain. *Ubr1* is located in the middle of mouse chromosome 2 and in the syntenic 15q15-q21.1 region of human chromosome 15. Mouse *Ubr1* spans ≈120 kilobases of genomic DNA and contains ≈50 exons. *Ubr1* is ubiquitously expressed in adults, with skeletal muscle and heart being the sites of highest expression. In mouse embryos, the *Ubr1* expression is highest in the branchial arches and in the tail and limb buds. The cloning of *Ubr1* makes possible the construction of *Ubr1*-lacking mouse strains, a prerequisite for the functional understanding of the mammalian N-end rule pathway.**

A number of regulatory circuits involve metabolically unstable proteins. Short *in vivo* half-lives are also characteristic of damaged or otherwise abnormal proteins (1–4). Features of proteins that confer metabolic instability are called degradation signals, or degrons. The essential component of one degradation signal, called the N-degron, is a destabilizing N-terminal residue of a protein (5, 6). The set of amino acid residues that are destabilizing in a given cell type yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar, but distinct, versions of the N-end rule pathway are present in all organisms examined, from mammals to fungi and bacteria (6–8).

In eukaryotes, the N-degron comprises two determinants: a destabilizing N-terminal residue and an internal lysine or lysines (8). The Lys residue is the site of formation of a multiubiquitin chain (9). The N-end rule pathway is thus one pathway of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, division, differentiation, and responses to stress (1, 3, 4, 10). In most of these processes, Ub acts through routes that involve the degradation of Ub-protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (11).

The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their enzymatic deamidation into the secondary destabilizing N-terminal residues Asp and Glu (12). The destabilizing activity of N-terminal Asp and Glu requires their enzymatic conjugation to Arg, one of the primary destabilizing residues (6). The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded N-recognin (also called E3), the recognition component of the N-end rule pathway (13). In *S. cerevisiae*, N-recognin is a 225-kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues—Phe, Leu, Trp, Tyr, Ile, Arg, Lys, and His. N-recognin has at least two substrate-binding sites. The type 1 site is specific for the basic N-terminal residues Arg, Lys, and His. The type 2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr, and Ile (6).

The known functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae* (through degradation of Cup9p, a transcriptional repressor of the peptide transporter Ptr2p); a role in controlling the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae*; the degradation of Gpa1p, a Gα protein of *S. cerevisiae*; and the degradation of alphaviral RNA polymerases in virus-infected metazoan cells (6, 14).

The mammalian counterpart of the yeast *UBR1*-encoded N-recognin (E3) was characterized biochemically in extracts from rabbit reticulocytes (15–17). Rabbit E3α was shown to be spe-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.



Abbreviations: Ub, ubiquitin; kb, kilobase; id., identity; si., similarity; BAC, bacterial artificial chromosome; FISH, fluorescence *in situ* hybridization; e*n*, embryonic day.

Data deposition: Nucleotide sequences reported in this work have been deposited in the GenBank database [accession nos. AF061555 (mouse *Ubr1* cDNA) and AF061556 (human *UBR1* cDNA)].

**To whom reprint requests should be addressed at: Division of Biology, 147-75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125. e-mail: avarsh@cco.caltech.edu.

††The names of mouse genes are in italics, with the first letter uppercase. The names of human and *S. cerevisiae* genes are also in italics, all uppercase. If human and mouse genes are named in the same sentence, the mouse gene notation is used. The names of *S. cerevisiae* proteins are Roman, with the first letter uppercase and an extra lowercase "p" at the end. The names of the corresponding mouse and human proteins are the same, except that all letters but the last "p" are uppercase. The latter usage is a modification of the existing convention (33), to facilitate simultaneous discussions of yeast, mouse, and human proteins. In some citations, the abbreviated name of a species precedes the gene's name.

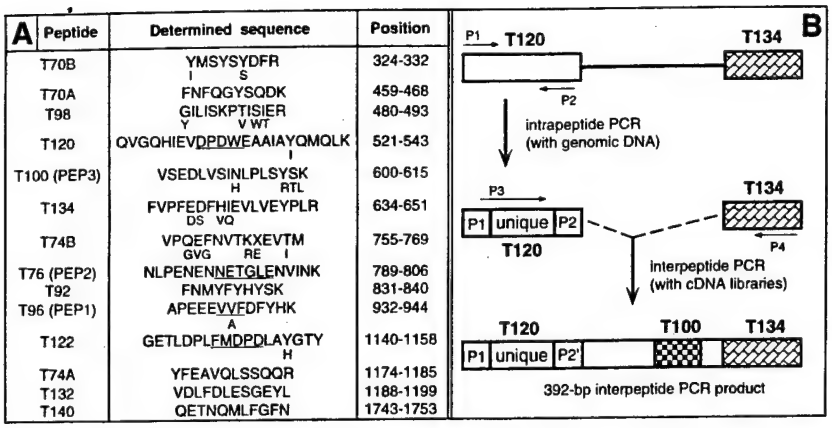



FIG. 1. Peptides of rabbit UBR1p (E3$\alpha$) and isolation of the mouse *Ubr1* cDNA. (*A*) Amino acid sequences of tryptic peptides of the purified rabbit UBR1p (see *Materials and Methods*). The alternative sets of peptide names, T-based and PEP1-PEP3 (in parentheses), refer to two different preparations of E3$\alpha$. The sequences of T120, T76, T96, and T122 that were encoded by DNA sequences identified through intrapeptide PCR are underlined. Residues deduced from the mouse *Ubr1* cDNA that differed from those inferred through peptide sequencing are indicated in a smaller font. The peptides' positions in the deduced sequence of mouse UBR1p are indicated. (*B*) The intrapeptide/interpeptide-PCR cloning strategy. The products of the initial intrapeptide PCR, derived from rabbit genomic DNA, were used to carry out interpeptide PCRwith a rabbit liver cDNA library (CLONTECH). The resulting 392-bp fragment of the rabbit *Ubr1* cDNA was used to isolate, using PCR and a λgt11 mouse liver cDNA library, the corresponding 392-bp mouse *Ubr1* cDNA fragment. This fragment then was used to screen the same cDNA library, yielding a 2.4-kb fragment of the mouse *Ubr1* cDNA that encoded several of the peptide-derived sequences of the rabbit UBR1p. The encoded sequence was also significantly similar to that of the N-terminal region of *S. cerevisiae* Ubr1p (13) and contained the putative start (ATG) codon of the mouse *Ubr1* ORF. To isolate the rest of the 5′ region of the *Ubr1* cDNA, 5′-rapid amplification of cDNA ends (RACE)–PCR (20) was performed with poly(A)$^+$ RNA from mouse L cells and a primer from the 2.4-kb DNA fragment. 3′-RACE–PCR (20) was used to amplify a downstream region of *Ubr1* cDNA. The resulting DNA fragment (nucleotides 2,470–3,467) then was used to screen a λgt10 mouse cDNA library from MEL-C19 cells. Five overlapping cDNA isolates (MR16, MR17, MR19, MR20, and MR23) that together spanned the entire *Ubr1* cDNA were mapped and subcloned into Bluescript II SK$^+$ (Stratagene), yielding the plasmid MR26, which contained the entire ORF of *Ubr1*. The ORF region of *Ubr1* cDNA was sequenced on both strands at least twice, using independently derived cDNA clones.

cifically required for the Ub-dependent degradation of proteins bearing either type 1 (basic) or type 2 (bulky hydrophobic) destabilizing N-terminal residues (7, 15, 16).

We began dissection of the mouse N-end rule pathway by isolating the *Ntan1* gene, which encodes the asparagine-specific N-terminal amidase (18, 19), a component of the mammalian N-end rule pathway, and by constructing mouse strains that lack *Ntan1* (Y.T.K. and A.V., unpublished data). Herein, we describe the cloning and characterization of the mouse and human cDNAs and genes†† that encode UBR1p (E3$\alpha$), a homolog of yeast Ubr1p and the main recognition component of the N-end rule pathway.

## MATERIALS AND METHODS

**Isolation and Partial Sequencing of Mammalian E3$\alpha$ (UBR1p).** Rabbit E3$\alpha$ was purified from reticulocyte extracts by using affinity chromatography with immobilized protein substrates of UBR1p and elution with dipeptides bearing destabilizing N-terminal residues (16). The resulting preparation was fractionated by SDS/PAGE. The band of ≈180-kDa E3$\alpha$ was excised and subjected to digestion with trypsin. Amino acid sequences were determined for 14 peptides of rabbit UBR1p (Fig. 1*A*) by using standard methods (20).

**Isolation of the Full-Length Mouse *Ubr1* cDNA.** A strategy that included the intrapeptide-interpeptide PCR (21) was used (see the legend to Fig. 1).

**Isolation of a Partial Human *UBR1* cDNA.** Poly(A)$^+$ RNA from human 293 cells was subjected to reverse transcription–PCR, using sets of primers corresponding to sequences of the mouse *Ubr1* cDNA. One of the reactions yielded a 1.0-kilobase (kb) fragment that encompassed a region of the human *UBR1* cDNA (Fig. 2).

**Mouse and Human Genomic *Ubr1* Fragments.** A library of mouse genomic DNA fragments (strain SvJ) in bacterial artificial chromosome (BAC) (22) vector (Genome Systems, St. Louis) was used, as described in the legend to Fig. 2.

**Northern, Southern, and Whole-Mount *in Situ* Hybridizations.** Mouse and human multiple-tissue Northern blots

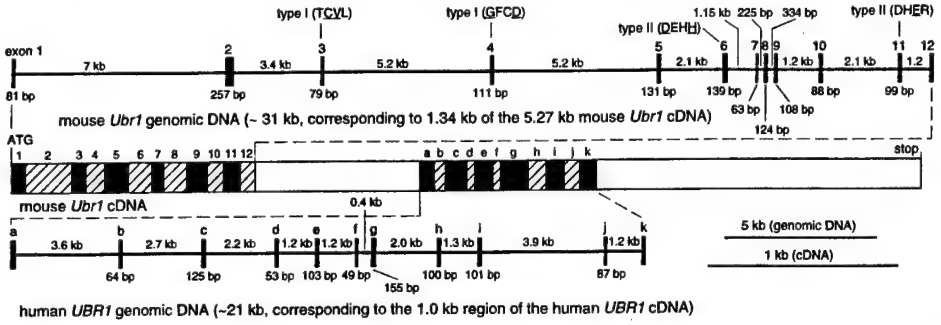



FIG. 2. The mouse and human *Ubr1* cDNAs and genes. Thick horizontal lines represent genomic DNA. The upper one is a ≈31-kb fragment of the mouse *Ubr1* gene that corresponds to a 1.34-kb region of mouse *Ubr1* cDNA (nucleotides 115–1,454). Vertical rectangles represent exons. Their lengths, and the lengths of the introns, are indicated, respectively, below and above the horizontal line. In a composite diagram of the *Ubr1* cDNA, the exons are depicted as alternatively shaded rectangles. For exon 1, only its translated region is indicated. Shown belowthe cDNA diagram is a ≈21-kb fragment of the human *UBR1* gene, corresponding to 1.0 kb of the indicated region of the human *UBR1* cDNA (nucleotides 2,218–3,227 of the mouse *Ubr1* cDNA sequence). The mouse and human *Ubr1* exons are denoted, respectively, by numbers and letters. Also indicated are the exon locations of some of the type 1 and type 2 substrate-binding sites of N-recognin (the essential amino acid residues are underlined) (A. Webster, M. Ghislain, and A.V., unpublished data; see the main text). Not shown are the 114-bp 5′-untranslated region (UTR) and the 1,010 bp 3′-UTR of the mouse *Ubr1* cDNA. To isolate mouse *Ubr1*, a library of mouse genomic DNA fragments in a BAC vector (see *Materials and Methods*) was screened with a fragment of the mouse *Ubr1* cDNA (nucleotides 105–1,333) as a probe, yielding seven BAC clones, of which BAC3 and BAC4 contained the entire *Ubr1* gene. The exon/intron organization of the first 31 kb (≈1/4) of the mouse *Ubr1* gene was determined by using exon-specific PCR primers to produce ≈40 genomic DNA fragments of the BAC3 insert that ranged in size from 1.3 to 18 kb. Regions encompassing the exon/intron junctions then were sequenced by using intron-specific primers. Fragments of the human genomic *UBR1* DNA were isolated by using primers derived from the 1.0-kb fragment of the human *UBR1* cDNA, the Expand High Fidelity PCR System (Roche Molecular Biochemicals, Indianapolis, IN), and genomic DNA from human 293 cells. The resulting four fragments were subcloned into pCR2.1 (Invitrogen), yielding the plasmids HR8, HR6–4, HR2–25, and HR7–2, whose partially overlapping inserts encompassed ≈21 kb of the human *UBR1* gene. Partial sequencing of the mouse and human genomic *Ubr1* fragments (≈20 kb of sequenced DNA) included all of the exon/intron junctions in these regions of *Ubr1*.

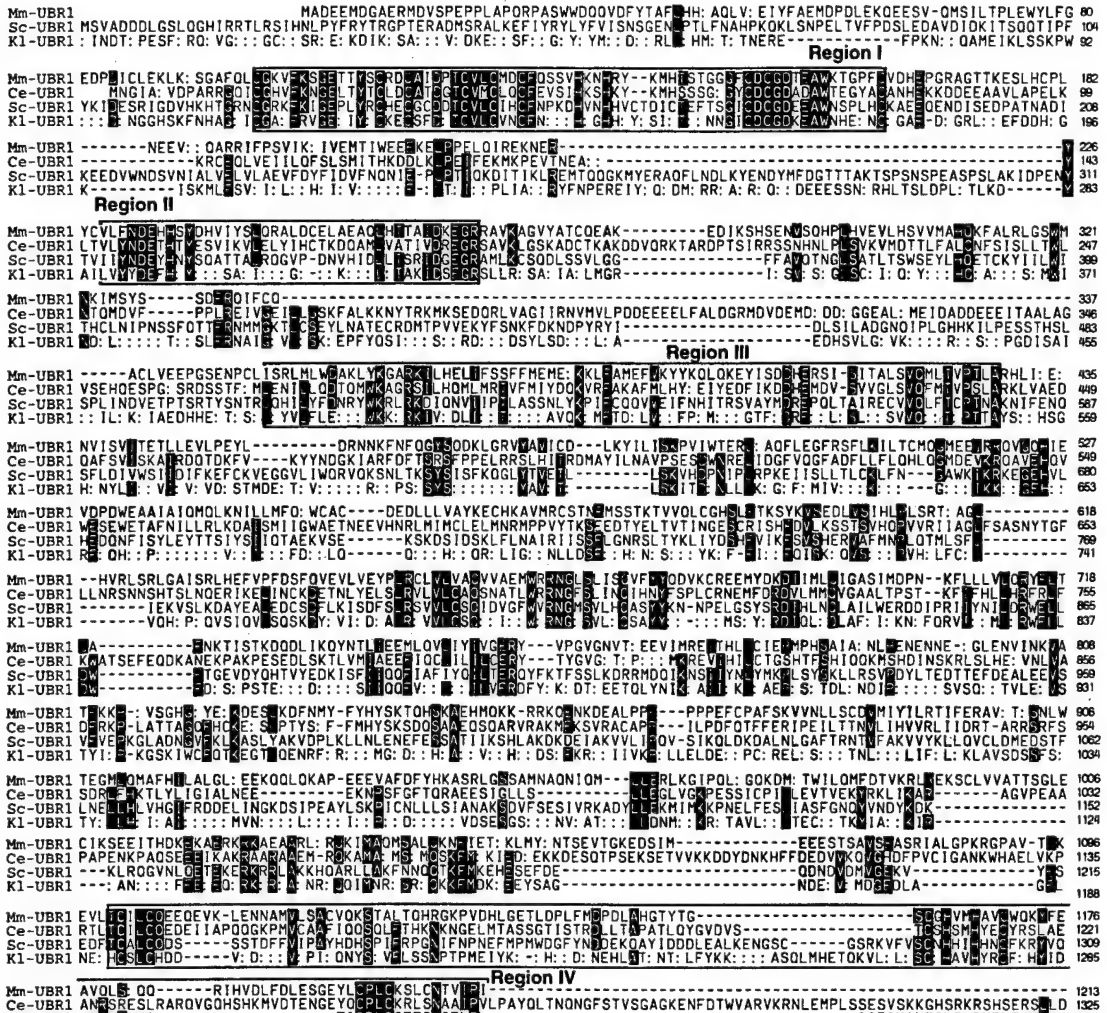



FIG. 3. Comparison of the deduced amino acid sequence of mouse UBR1p (Mm-UBR1) with those of *C. elegans* UBR1p (Ce-UBR1), *S. cerevisiae* Ubr1p (Sc-UBR1), and *K. lactis* Ubr1p (Kl-UBR1). White-on-black and gray shadings highlight, respectively, identical and similar residues. The residues of UBR proteins that are identical to those of *S. cerevisiae* Ubr1p are denoted by double dots, at positions where the identity involves just one non-*cerevisiae* protein. Also indicated are the regions of significant similarity among the four proteins. *K. lactis UBR1* was cloned through its crosshybridization to *S. cerevisiae UBR1* (P. Waller and A.V., unpublished data).

(CLONTECH), and either mouse or human *Ubr1* cDNA fragments labeled with $^{32}P$ were used (20). Southern hybridizations were carried out by using standard techniques (20). Mouse embryos were staged, fixed, and processed for *in situ* hybridization as described (23). For sectioning, the stained embryos were embedded in OCT medium (Sakura Finetek, Torrance, CA). A 1.2-kb *Ubr1* cDNA fragment (nucleotides 3,150–3,355) was used as a template for synthesizing antisense- or sense-strand RNA probes labeled with digoxigenin (23).

**Chromosome Mapping of the Mouse and Human *Ubr1*.** The mapping of mouse *Ubr1* was carried out by using the interspecific backcross analysis (24), essentially as described (18). Human *UBR1* was mapped by using fluorescence *in situ* hybridization

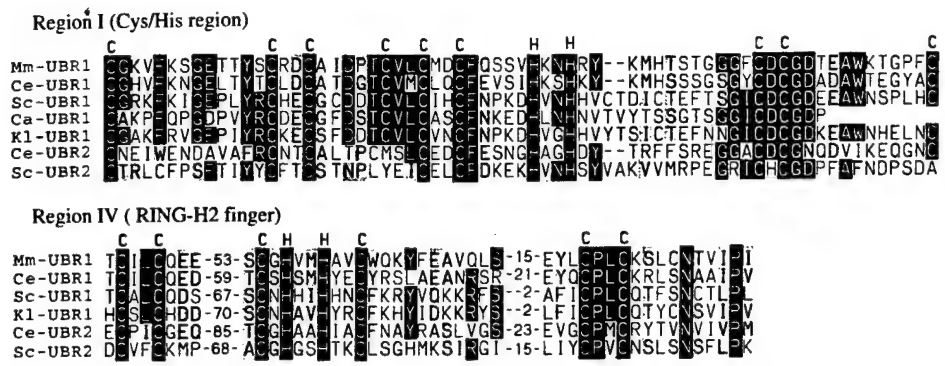



FIG. 4. Two Cys/His domains of the UBR protein family. Comparison of the putative zinc finger (region I) and RING-H2 finger (region IV) with the corresponding sequences from the other species in Fig. 3, and also with *C. albicans* Ubr1p (Ca-UBR1), *C. elegans* UBR2p (Ce-UBR2), and *S. cerevisiae* Ubr2p (Sc-UBR2). Numbers indicate the lengths of gaps. The conserved Cys and His residues are indicated.

(FISH) with mitotic chromosomes from human lymphocytes (25). The probe was a mixture of the HR8, HR6–4, HR2–25, and HR7–2 plasmids, labeled with biotin using biotinylated dATP and the BioNick labeling kit (Life Technologies, Grand Island, NY), and detected by using fluorescein isothiocyanate-avidin.

## RESULTS AND DISCUSSION

**Isolation of the Mouse *Ubr1* cDNA.** Tryptic peptides of the purified rabbit E3α (UBR1p) (16) were isolated and sequenced (see *Materials and Methods*), yielding 14 short regions of E3α (Fig. 1*A*). These regions lacked significant similarities to the deduced sequence of *S. cerevisiae* Ubr1p (13). We used intrapeptide PCR (21) to identify a unique (nondegenerate) sequence of the rabbit *Ubr1* cDNA. This method allows amplification of a short unique DNA sequence by using two degenerate PCR primers (derived by reverse translation) that flank this sequence and correspond to the outermost regions of a single peptide (Fig. 1*B*). Several intrapeptide nucleotide sequences were obtained this way (Fig. 1). These sequences, together with those of the original degenerate primers, then were used to amplify a 392-bp fragment of the rabbit *Ubr1* cDNA, using interpeptide PCR (Fig. 1). This fragment encoded peptides T120 and T134 at either end and peptide T100 in the middle (Fig. 1*B*). A homologous 392-bp fragment of the mouse *Ubr1* cDNA then was amplified by using the same method (Fig. 1*B*). The rabbit and mouse 392-bp *Ubr1* cDNA fragments were 88% and 89% identical at the nucleotide and amino acid sequence levels, respectively, but lacked significant similarities to *S. cerevisiae UBR1* (data not shown).

The 392-bp mouse *Ubr1* cDNA fragment then was used, in conjunction with standard cDNA library screening and rapid amplification of cDNA ends–PCR (20), to isolate multiple *Ubr1* cDNA fragments, and to assemble them into a 5,271-bp ORF encoding a 1,757-residue protein (pI of 6.0), whose size, 200 kDa, was close to the estimated size of the isolated rabbit UBR1p (E3α), ≈180 kDa (16) (Figs. 2 and 3). The inferred ATG start codon (Fig. 2), within the sequence CTTAAGATGGCG, is preceded by two in-frame stop codons, at positions −48 and −93, and is located in a favorable Kozak context (26), with A and G at positions −3 and +4, respectively. There are two more ATGs, five and 11 codons downstream of the inferred one. These alternative start codons are in a favorable Kozak context as well.

**Cloning and Partial Characterization of the Mouse and Human *Ubr1* Genes.** A fragment of the mouse *Ubr1* cDNA was used to isolate a ≈120-kb mouse *Ubr1* genomic DNA clone, carried in a BAC vector (22). We determined the exon/intron organization and restriction map of the ≈31-kb region of *Ubr1* that corresponded to the 1,340-bp 5′-region of the mouse *Ubr1* cDNA (nucleotides 105–1,333) (Fig. 2). The lengths of the 12 exons in this region of mouse *Ubr1* range from 63 to 257 bp (Fig. 2).

The nucleotide and deduced amino acid sequences of the 1.0-kb human *UBR1* cDNA fragment (see *Materials and Methods*), located approximately in the middle of *UBR1* cDNA (nucleotides 2,218–3,227 of the mouse *Ubr1* cDNA) (Fig. 2), were, respectively, 91% and 94% identical to the corresponding mouse *Ubr1* cDNA and UBR1p sequences. Overlapping genomic DNA fragments of human *UBR1* that, together, encompassed a ≈21-kb region of the human *UBR1* gene and corresponded to the 1.0-kb fragment of the human *UBR1* cDNA (Fig. 2), were isolated from human DNA by using cDNA-derived primers and PCR. Partial sequencing showed that this ≈21-kb region of human *UBR1* contained 11 exons whose length ranged from 49 to 155 bp, a distribution of exon lengths similar to that in a different region of mouse *Ubr1* (Fig. 2). All of the sequenced exon/intron junctions (≈23 exons), which encompassed a ≈52-kb region of the mouse and human *Ubr1*, contained the consensus GT and AG dinucleotides characteristic of the mammalian nuclear pre-mRNA splice sites (data not shown) (20). Extrapolating from these data on the mouse and human *Ubr1* genes and the corresponding regions of their cDNAs (Fig. 2), a mammalian *Ubr1* gene is expected to be ≈120 kb long and to contain ≈50 exons.

**The Mouse UBR1p Protein and its Homologs.** The low overall sequence similarity of mouse UBR1p (E3α) to Ubr1p of either *S. cerevisiae* [22% identity (id.), 48% similarity (si.)] or another budding yeast, *Kluyveromyces lactis* (21% id., 48% si.), belied the presence of five regions, denoted I–V, which were significantly similar between the mouse and yeast versions of UBR1p (Figs. 3 and 4). By contrast, the *Ntan1*-encoded asparagine-specific N-terminal amidase, the most upstream component of the mouse N-end rule pathway, lacks sequence similarities to its *S. cerevisiae* counterpart Nta1p (12, 18). Database searches identified other likely homologs of mouse UBR1p, in particular the 1,927-residue protein of the nematode *Caenorhabditis elegans* (GenBank accession no. U88308) (32% id., 53% si.; termed *Ce-Ubr1*); the 1,872-residue *S. cerevisiae* protein (GenBank accession no. Z73196) (21% id., 47% si.; termed *Sc-UBR2*; ref. 4); the 2,168-residue *C. elegans* protein (GenBank accession no. U40029) (21% id., 45% si.; termed *Ce-Ubr2*); and the 794-residue CER3p protein of the plant *Arabidopsis thaliana* (GenBank accession no. X95962) (26% id., 49% si.). CER3p is involved in wax biosynthesis in *A. thaliana* (27). In addition, a 147-residue sequence of the yeast *Candida albicans* (http://alces.med.umn.edu/bin/genelist?LUBR1) was similar to the N-terminal region of mouse UBR1p (Fig. 4).

The presence of high-similarity regions I–V among these deduced sequences (Figs. 3 and 4) suggested the existence of a distinct protein family, termed UBR. The 66-residue region I, near the N terminus of UBR1p, is a particularly clear UBR family-identifying region (e.g., 61% id., 75% si. between mouse and *C. elegans* UBR1p) (Figs. 3 and 4).

Recent genetic analyses of *S. cerevisiae* Ubr1p (N-recognin) have shown that the regions I–III contain residues essential for the recognition of N-end rule substrates by Ubr1p. In particular, Cys-145, Val-146, Gly-173, and Asp-176 of region I were identified as essential residues of the type 1 binding site of *S. cerevisiae* Ubr1p (A. Webster, M. Ghislain, and A.V., unpublished data). All four of these residues were conserved between the yeast, mouse, and *C. elegans* UBR1p (Figs. 3 and 4). Region I is present in all of the known UBR family members except CER3p of *A. thaliana*, which contains only regions IV and V (Fig. 4). Region I encompasses a Cys/His-rich domain, Cys-$X_{12}$-Cys-$X_2$-Cys-$X_5$-

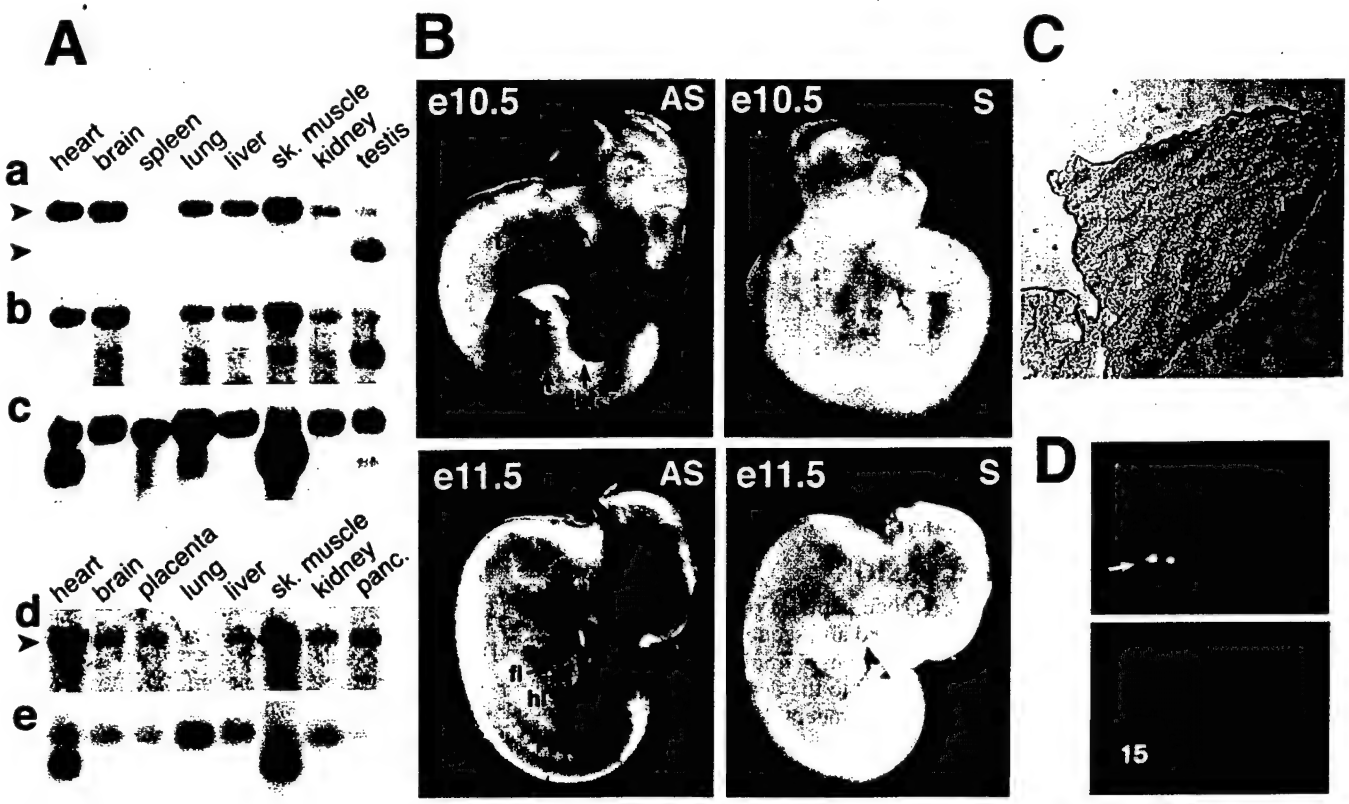



FIG. 5. Northern and *in situ* hybridizations with mouse and human *Ubr1*. (*A*) Membranes containing electrophoretically fractionated poly(A)$^+$ mRNA from different mouse (*a–c*) or human (*d* and *e*) tissues were hybridized with either a 2-kb 5′-proximal (nucleotides 116–2,124) mouse *Ubr1* cDNA fragment (*a*), its 0.64-kb 3′-proximal (nucleotides 4,749–5,388) fragment (*b*), a 1-kb human *UBR1* cDNA fragment (*d*), or the human β-actin cDNA fragment (*c* and *e*). The upper arrows in *a* and *d* indicate the ≈8-kb *Ubr1* transcript. The lower arrow in *a* indicates the ≈6-kb testis-specific *Ubr1* transcript. In the RNA sample from mouse spleen, the *Ubr1* transcript (but not the actin transcript) may have been degraded (*a–c*). (*B*) Expression of *Ubr1* in e10.5 and e11.5 mouse embryos. Whole-mount *in situ* hybridization was carried out with either antisense (AS) or sense (S, negative control) *Ubr1* cDNA probes (see *Materials and Methods*). The regions of high *Ubr1* expression are indicated by arrows (t, tail; fl, forelimb buds; hl, hindlimb buds). The branchial arches, where *Ubr1* is also highly expressed in e10.5 embryos (data not shown), are not visible in this e10.5 embryo. (*C*) Expression of *Ubr1* in the surface ectoderm of limb buds. Shown is a transverse section of a forelimb bud of an e10.5 embryo (se, surface ectoderm). (*D*) FISH analysis of human *UBR1*. (*Upper*) An example of the *UBR1*-specific FISH signal (arrow). (*Lower*) The same mitotic spread stained with 4′-6-diamino-2-phenylindole (DAPI) to visualize the chromosomes (see also Fig. 6).

Cys-$X_2$-Cys-$X_2$-Cys-$X_5$-His-$X_2$-His-X($_{12-14}$)-Cys-$X_1$-Cys-$X_{11}$-Cys (Figs. 3 and 4), which is distinct from the known consensus sequences of zinc fingers and other Cys/His-motifs. Residues Asp-318, His-321, and Glu-560 of *S. cerevisiae* Ubr1p, which have been identified as essential for the type 2 binding site of this N-recognin (A. Webster, M. Ghislain, and A.V., unpublished data), were found to be retained in region II (Asp-318 and His-321) and region III (Glu-560) of the mouse and *C. elegans* UBR1p (Fig. 3).

Region IV contains another Cys/His-rich domain of UBR1p, Cys-$X_2$-Cys-*loop 1*-Cys-$X_1$-His-$X_2$-His-$X_2$-Cys-*loop 2*-Cys-$X_2$-Cys (Figs. 3 and 4), which is present in all of the UBR family members, and fits the consensus sequence of the RING-H2 finger, a subfamily of the previously defined RING motif (28). At least some of the RING-H2 sequences are sites of specific protein–protein interactions (28). Apc11p, a subunit of the Ub-protein ligase complex called the cyclosome (2) or the anaphase promoting complex, also contains a RING-H2 finger (29).

Another area of similarity (24–50% id., 46–70% si.) among the UBR family members is region V (Fig. 3 and data not shown). This region, 115 residues long in mouse UBR1p, near the protein's C terminus, is particularly similar between mouse and *C. elegans* UBR1p (50% id., 70% si.) (Fig. 3). Region V is located 4–14 residues from the UBR proteins' C termini, the exceptions being the *S. cerevisiae* and *K. lactis* Ubr1p, which bear, respectively, 132- and 159-residue tails of unknown function that are rich in the acidic Asp/Glu residues (36% and 33%) (Fig. 3). No significant similarities could be detected between mammalian UBR1p and other E3s (recognins) of the metazoan Ub system, including E6AP (30) and subunits of the cyclosome/anaphase promoting complex, except for the presence of a RING-H2 finger domain in the latter (29). [Different E3 proteins of the Ub system recognize different degrons in protein substrates, thereby defining distinct Ub-dependent proteolytic pathways (1, 4).]

**Expression of Mouse and Human *Ubr1*.** The 5′- and 3′-proximal mouse cDNA probes yielded similar results, detecting a single ≈8-kb transcript in several tissues (Fig. 5*Aa* and *Ab*). In the testis, however, the ≈8-kb species of *Ubr1* mRNA was a minor one, the major species being ≈6 kb (Fig. 5*Aa*). The levels of either mouse or human *Ubr1* mRNA were highest in skeletal muscle and heart (Fig. 5*A*). The expression of mRNA encoding $E2_{14K}$, one of the mouse Ub-conjugating (E2) enzymes and a likely component of the mouse N-end rule pathway (6), was also highest in skeletal muscle and heart (18).

The distinct *Ubr1* mRNA pattern in the testis (Fig. 5*A*) was reminiscent of the analogous expression pattern of *Ntan1* mRNA, which encodes the Asn-specific N-terminal amidase, another component of the mammalian N-end rule pathway. Specifically, the size of the major species of *Ntan1* mRNA was ≈1.4 kb in all of the examined mouse tissues except testis, where the major species was ≈1.1 kb (18). The ≈1.1-kb *Ntan1* transcript recently was found to hybridize only to the 3′-half (exons 6–10 but not exons 1–5) of the *Ntan1* ORF (Y.T.K. and A.V., unpublished data). The functional significance of the testis-specific *Ubr1* and *Ntan1* expression patterns remains to be understood.

We used whole-mount *in situ* hybridization to examine the expression of *Ubr1* during embryogenesis. In e9.5 (9.5 days old)

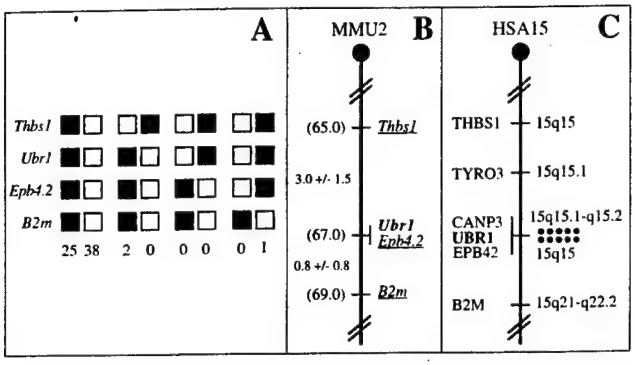



FIG. 6. Chromosomal locations of the mouse and human *Ubr1* genes. (*A*) Mouse *Ubr1* was mapped to the middle of mouse chromosome 2 by using interspecific (*M. musculus-M. spretus*) backcross analysis (18, 24). Shown are the segregation patterns of mouse *Ubr1* and the flanking genes in 66 backcross animals that were typed for all loci. For individual pairs of loci, more than 66 animals were typed. Each column represents the chromosome identified in the backcross progeny that was inherited from the [*M. musculus* C57BL/6J × *M. spretus*] $F_1$ parent. Filled and empty squares represent, respectively, C57BL/6J and *M. spretus* alleles. The numbers of offspring that inherited each type of chromosome 2 are listed below the columns. (*B*) A partial mouse chromosome 2 linkage map (MMU2), showing *Ubr1* in relation to the linked genes *Thbs1*, *Ebp4.2*, and *B2m*, and also, on the left, the corresponding recombination distances between the loci, in centimorgans, and the map locations, in parentheses. (*C*) A partial human chromosome 15 linkage map (HSA15). Each dot on the right, in the 15q15-q21.1 region, corresponds to the actually observed *UBR1*-specific double-dot FISH signal detected on human chromosome 15 (see also Fig. 5*D*).

mouse embryos, the expression of *Ubr1* was highest in the branchial arches and in the buds of forelimbs and the tail (data not shown). In e10.5 embryos, the expression of *Ubr1* became high in the hindlimb buds as well (Fig. 5*B*). This pattern was maintained in the limb buds of e11.5 embryos (Fig. 5*B*). High expression of *Ubr1* in the limb buds was confined predominantly to the surface ectoderm (Fig. 5*C*). This pattern of *Ubr1* expression in embryos (Fig. 5 *B* and *C*) is similar, if not identical, to that of *Ntan1*, which encodes asparagine-specific N-terminal amidase (18) (Y.T.K. and A.V., unpublished data), consistent with UBR1p and NTAN1p being components of the same pathway.

The enhanced expression of *Ubr1* in the embryonic limb buds (Fig. 5 *B* and *C*) is interesting in view of the conjecture that the N-end rule pathway might be required for limb regeneration in amphibians (31). The injection of dipeptides bearing destabilizing N-terminal residues into the stumps of amputated forelimbs of the newt was observed to delay limb regeneration, whereas the injection of dipeptides bearing stabilizing N-terminal residues had no effect (31). Rigorous tests of this and other suggested functions of the metazoan N-end rule pathway (6) will require mouse strains that lack *Ubr1*.

**Chromosome Mapping of Mouse and Human *Ubr1*.** The chromosomal location of mouse *Ubr1* was determined by interspecific backcross analysis, using DNA derived from matings of [(C57BL/6J × *Mus spretus*)$F_1$ × C57BL/6J] mice (Fig. 6 *A* and *B*) (18, 24). Mouse *Ubr1* is located in the central region of chromosome 2 and is linked to the *Thbs1*, *Epb4.2*, and *B2m* genes, the most likely gene order being centromere-*Thbs1*-*Ubr1*-*Epb4.2*-*B2m* (Fig. 6*B* and data not shown).

The chromosomal location of human *UBR1* was determined by using FISH (25), with human *UBR1* genomic DNA fragments as probes (Figs. 5*D* and 6*C*). This mapping placed *UBR1* at the 15q15–15q21.1 region of the human chromosome 15, an area syntenic with the independently mapped position of mouse *Ubr1* (Fig. 6). *Ubr1* is located in the regions of human chromosome 15 and mouse chromosome 2 that appear to be devoid of the previously mapped but uncloned mutations. Mutations in the human gene *CANP3*, which encodes a subunit of calpain and is located very close, if not adjacent, to *UBR1*, have been shown to cause a myopathy called the limb-girdle muscular distrophy (32).

**Concluding Remarks.** Isolation of the mouse and human *Ubr1* cDNAs and genes (Figs. 2–6) should enable functional understanding of the mammalian N-end rule pathway, in part through the construction and analysis of mouse strains that lack *Ubr1*. Recent searches in GenBank identified several mouse and human sequences in expressed sequence tag databases that exhibited significant similarity to the C-terminal region of mouse UBR1p. The cloning and characterization of the corresponding cDNAs have shown that there exist at least two distinct mouse (and human) genes, termed *Ubr2* and *Ubr3*, which encode proteins that are significantly similar to mouse UBR1p (Y.T.K. and A.V., unpublished data). Molecular and functional analyses of these *Ubr1* homologs are under way.

We are grateful to A. Webster and M. Ghislain for permission to cite their unpublished data. We thank members of the Varshavsky lab, especially I. V. Davydov, for helpful discussions, and L. Peck, G. Turner, H. Rao, A. Kashina, and F. Du for comments on the manuscript. Y.T.K. thanks B. Yu for sharing his Northern hybridization data on human β-actin mRNA. We gratefully acknowledge the sequencing of *K. lactis UBR1* by P. Waller. N.G.C. and N.A.J. thank D. J. Gilbert and D. B. Householder for excellent technical assistance. D.K.G. was a Scholar of the Leukemia Society of America. This study was supported by National Institutes of Health grants to A.V. (DK39520 and GM31530), V.A.F. (NS29542), and D.K.G. (GM45314), and by a grant to N.G.C. from the National Cancer Institute.

1. Varshavsky, A. (1997) *Trends Biochem. Sci.* **22,** 383–387.
2. Hershko, A. (1997) *Curr. Opin. Cell. Biol.* **9,** 788–799.
3. Haas, A. J. & Siepman, T. J. (1997) *FASEB J.* **11,** 1257–1268.
4. Hochstrasser, M. (1996) *Annu. Rev. Genet.* **30,** 405–439.
5. Bachmair, A., Finley, D. & Varshavsky, A. (1986) *Science* **234,** 179–186.
6. Varshavsky, A. (1997) *Genes Cells* **2,** 13–28.
7. Gonda, D. K., Bachmair, A., Wünning, I., Tobias, J. W., Lane, W. S. & Varshavsky, A. (1989) *J. Biol. Chem.* **264,** 16700–16712.
8. Bachmair, A. & Varshavsky, A. (1989) *Cell* **56,** 1019–1032.
9. Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K. & Varshavsky, A. (1989) *Science* **243,** 1576–1583.
10. Pickart, C. M. (1997) *FASEB J.* **11,** 1055–1066.
11. Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. (1998) *Cell* **92,** 367–380.
12. Baker, R. T. & Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 12065–12074.
13. Bartel, B., Wünning, I. & Varshavsky, A. (1990) *EMBO J.* **9,** 3179–3189.
14. Byrd, C., Turner, G. C. & Varshavsky, A. (1998) *EMBO J.* **17,** 269–277.
15. Reiss, Y., Kaim, D. & Hershko, A. (1988) *J. Biol. Chem.* **263,** 2693–269.
16. Reiss, Y. & Hershko, A. (1990) *J. Biol. Chem.* **265,** 3685–3690.
17. Hershko, A. & Ciechanover, A. (1992) *Annu. Rev. Biochem.* **61,** 761–807.
18. Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N. A., Copeland, N. J. & Varshavsky, A. (1996) *J. Biol. Chem.* **271,** 28521–28532.
19. Stewart, A. E., Arfin, S. M. & Bradshaw, R. A. (1995) *J. Biol. Chem.* **270,** 25–28.
20. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G. & Struhl, K. (1996) *Current Protocols in Molecular Biology* (Wiley Interscience, New York).
21. Bredt, D. S., Hwang, P. M., Glatt, C. E., Lowenstein, C., Reed, R. R. & Snyder, S. H. (1991) *Nature (London)* **351,** 714–718.
22. Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y. & Simon, M. I. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8794–8797.
23. Conlon, R. A. & Rossant, J. (1992) *Development (Cambridge, U.K.)* **116,** 357–368.
24. Copeland, N. G. & Jenkins, N. A. (1991) *Trends Genet.* **7,** 113–118.
25. Dracopoli, N. C., Haines, J. L., Korf, B. R., Moir, T. D., Morton, C. C., Seidman, C. E., Seidman, J. G. & Smith, D. R. (1994) *Current Protocols in Human Genetics* (Wiley Interscience, New York).
26. Kozak, M. (1996) *Mamm. Genome* **7,** 563–574.
27. Hannoufa, A., Negruk, V., Eisner, G. & Lemieux, B. (1996) *Plant J.* **10,** 459–467.
28. Borden, K. L. & Freemont, P. S. (1996) *Curr. Opin. Struct. Biol.* **6,** 395–401.
29. Yu, H., Peters, J. M., King, R. W., Page, A. M., Hieter, P. & Kirschner, M. W. (1998) *Science* **279,** 1219–1222.
30. Huibregtse, J. M., Scheffner, M., Beaudenon, S. & Howley, P. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 2563–2567.
31. Taban, C. H., Hondermarck, H., Bradshaw, R. A. & Boilly, B. (1996) *Experientia* **52,** 865–870.
32. Richard, I., Broux, O., Allamand, V., Fougerousse, F., Chiannilkulchai, N., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C., *et al.* (1995) *Cell* **81,** 27–40.
33. Stewart, A. (1995) *Trends in Genetics Nomenclature Guide* (Elsevier, Cambridge, U.K.).

ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS
Vol. 357, No. 2, September 15, pp. 317–325, 1998
Article No. BB980829

# The N-End Rule Pathway in *Xenopus* Egg Extracts

Ilia V. Davydov, Debabrata Patra, and Alexander Varshavsky[1]
*Division of Biology, California Institute of Technology, Pasadena, California 91125*

Received April 7, 1998, and in revised form June 30, 1998

**Ubiquitin-dependent degradation of intracellular proteins underlies a multitude of biological processes, including the cell cycle, cell differentiation, and responses to stress. One ubiquitin-dependent proteolytic system is the N-end rule pathway, whose targets include proteins that bear destabilizing N-terminal residues. This pathway, which has been characterized only in somatic cells, is shown here to be present also in germ line cells such as the eggs of the amphibian *Xenopus laevis*. We demonstrate that the set of destabilizing residues in the N-end rule pathway of *Xenopus* eggs is similar, if not identical, to that of somatic cells such as mammalian reticulocytes and fibroblasts. It is also shown that the degradation of engineered N-end rule substrates in egg extracts can be strongly and selectively inhibited by dipeptides bearing destabilizing N-terminal residues. This result allowed us to ask whether selective inhibition of the N-end rule pathway in egg extracts influences the apoptosis-like changes that are observed in these extracts. A dipeptide bearing a bulky hydrophobic (type 2) destabilizing N-terminal residue was found to delay the apoptotic changes in egg extracts, whereas dipeptides bearing basic (type 1) destabilizing N-terminal residues had no effect. High activity of the N-end rule pathway in egg extracts provides an alternative to reticulocyte extracts for the *in vitro* analyses of this pathway.**


*Key Words:* **ubiquitin; proteolysis; N-end rule; apoptosis; germ line; *Xenopus*.**

A number of regulatory circuits, including those that control the cell cycle, cell differentiation, and responses to stress, involve metabolically unstable proteins (1–4). A short *in vivo* half-life of a regulator provides a way to generate its spatial gradients and allows for rapid adjustments of its concentration, or subunit composition, through changes in the rate of its synthesis or degradation. Damaged or otherwise abnormal proteins tend to be short-lived as well.

Features of proteins that confer metabolic instability are called degradation signals, or degrons (5). The essential component of one degradation signal, called the N-degron, is a destabilizing N-terminal residue of a protein (6, 7). The set of amino acid residues which are destabilizing in a given cell yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct versions of the N-end rule pathway are present in all organisms examined, from mammals to fungi and bacteria (7).

In eukaryotes, an N-degron comprises at least two determinants: a destabilizing N-terminal residue and an internal lysine or lysines (8–10). The Lys residue is the site of formation of a multiubiquitin chain (11). The N-end rule pathway is thus a part of the ubiquitin (Ub)[2] system. Ub is a 76-residue eukaryotic protein whose covalent conjugation to other proteins plays a role in a multitude of biological processes (1, 12–15). In most of them, Ub acts through routes that involve the degradation of Ub-protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (16–18).

It was found that linear Ub fusions are rapidly cleaved by Ub-specific proteases (UBPs) at the Ub-protein junction, making possible the production of otherwise identical proteins bearing different N-terminal residues. This technical advance led to the discovery of the N-end rule pathway (6, 7). The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are *tertiary* destabilizing N-terminal residues in that they function through their conversion, by enzymatic deamidation, into the *second-*

[1] To whom correspondence should be addressed at Division of Biology, 147-75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125. Fax: (626) 440-9821. E-mail: avarsh@cco.caltech.edu.

[2] Abbreviations used: Ub, ubiquitin; DHFR, mouse dihydrofolate reductase; R-transferase, Arg-tRNA-protein transferase; CSF, cytostatic factor; TCA, trichloroacetic acid; DTT, dithiothreitol; AMC, 7-amino-methylcoumarin; UBP, Ub-specific processing protease.

0003-9861/98 $25.00

*ary* destabilizing N-terminal residues Asp and Glu. The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by Arg-tRNA-protein transferase (R-transferase), to Arg, one of the *primary* destabilizing residues (19–21). The primary destabilizing N-terminal residues are bound directly by N-recognin (also called E3), the recognition component of the N-end rule pathway. In *S. cerevisiae,* N-recognin is a 225-kDa protein, encoded by *UBR1,* that targets potential N-end rule substrates through the binding to their primary destabilizing N-terminal residues — Phe, Leu, Trp, Tyr, Ile, Arg, Lys, or His (7, 22). N-recognin has at least two substrate-binding sites. The type 1 site is specific for the basic N-terminal residues Arg, Lys, and His. The type 2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr, and Ile (7).

The currently known physiological substrates of the N-end rule pathway are the RNA polymerase of Sindbis virus (23), the *GPA1*-encoded G$\alpha$ subunit of the *S. cerevisiae* heterotrimeric G protein (24), and Cup9p, a short-lived transcriptional repressor in *S. cerevisiae* that controls the expression of *PTR2,* which encodes a peptide transporter (25). The Ubr1p-Cup9p-Ptr2p circuit, which controls the import of peptides in yeast, is the first clear instance of a physiological function of the N-end rule pathway (25). Among the cells of multicellular organisms, this proteolytic system was characterized in rabbit reticulocytes (19) and L cells, a line of fibroblast-like mouse cells (26). Recently, it was suggested that the N-end rule pathway might play a role in apoptosis (programmed cell death) (7). A way to verify this conjecture would be to use a mutant that lacks the N-end rule pathway. Such mutants were constructed and characterized in *Escherichia coli* and *S. cerevisiae* (7) but not yet in multicellular organisms. Therefore we considered the use of N-end rule inhibitors.

Previous work has shown that the addition of amino acid derivatives such as dipeptides that bear destabilizing N-terminal residues to reticulocyte extract results in a strong and selective inhibition of the N-end rule pathway in the extract (19, 27). Specifically, dipeptides bearing type 1 destabilizing N-terminal residues inhibited the degradation of test N-end rule substrates bearing basic (type 1) destabilizing N-terminal residues but had no effect on the degradation of substrates bearing type 2 N-terminal residues. A converse pattern was observed with dipeptides bearing bulky hydrophobic (type 2) destabilizing N-terminal residues (19). Dipeptides added to *S. cerevisiae* cultures have been demonstrated to inhibit the N-end rule pathway *in vivo* as well (28). However, experiments with the N-end rule pathway of intact mammalian cells have shown dipeptides to be ineffective inhibitors in this setting[3].

[3] F. Lévy and A. Varshavsky, unpublished data.

Given this latter constraint, we sought to explore the physiological consequences of inhibiting a metazoan N-end rule pathway in the inhibitor-accessible setting of a cell-free system. We also wished to determine whether the N-end rule pathway is present in germ line cells such as the eggs of the amphibian *Xenopus laevis,* a major experimental organism in studies of embryogenesis and cell cycle control (4, 29). Our findings are described below.

## MATERIALS AND METHODS

*Construction of plasmids.* The plasmids pT7-UbMe$^K$DHFRhis and pT7-UbRe$^K$DHFRhis expressed Ub-Met-e$^K$-DHFR-His$_6$ and Ub-Arg-e$^K$-DHFR-His$_6$ fusions (denoted, respectively, as Ub-Met-DHFR and Ub-Arg-DHFR) in *E. coli* from the T7 polymerase promoter (see Results for the definitions of e$^K$ and other terms). These plasmids were constructed from pEJJ1-M and pEJJ1-R (30). The *Kpn*I-*Hin*dIII fragment of pEJJ1-M and pEJJ1-R was replaced with a synthetic oligonucleotide duplex (5′-CCATCACCATCACCATCACTAAA-3′ and 5′-AGCTTTTAGTGATGGTGATGGTGATGGGTAC-3′) that encoded the His$_6$ tag and bore the *Kpn*I and *Hin*dIII overhangs. The other pT7-UbXe$^K$DHFRhis plasmids, which encoded otherwise identical Ub-X-DHFR fusions bearing different X residues (Met, Arg, Leu, Phe, Cys, Asp, or Asn) (Fig. 1), were constructed from pT7-UbRe$^K$DHFRhis by site-directed mutagenesis, using PCR (31).

*Overexpression, labeling, and purification of Ub-X-e$^K$-DHFR-His$_6$ proteins.* A pT7-UbXe$^K$DHFRhis plasmid was introduced into *E. coli* BL21 (DE3) (31). Protein expression was induced in *E. coli* by adding isopropyl-1-thio-$\beta$-*D*-galactopyranoside (IPTG), and cells were labeled with [$^{35}$S]methionine/cysteine ($^{35}$S-Express, New England Nuclear, Boston, MA), as described (30). The labeled cells were collected by centrifugation and disrupted by sonication, and a $^{35}$S-labeled Ub-X-e$^K$-DHFR-His$_6$ (Ub-X-DHFR) test protein was purified by affinity chromatography under nondenaturing conditions, using the Ni-NTA Spin Kit (Qiagen, Chatsworth, CA), and dialyzed against 1 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 0.1 M Tris–HCl (pH 7.7), frozen rapidly, and stored at −80°C in samples that were to be thawed only once. The specific radioactivity of [$^{35}$S]Ub-X-DHFR proteins was 5–10 × $10^5$ cpm/$\mu$g.

*Degradation assays in Xenopus egg extracts.* Cytostatic factor (CSF)-arrested egg extracts were prepared from *Xenopus* eggs as described (32). In most experiments, cycloheximide (0.1 mg/ml) was added to the extracts to preclude reincorporation of [$^{35}$S]methionine into newly made proteins. In some experiments (see the legends to figures), egg extracts were activated by the addition of 0.4 mM $CaCl_2$ 1 h before the assay. [$^{35}$S]Ub-X-DHFR test proteins were added to the extract to the final concentration of ~25 $\mu$g/ml. Dipeptides were added together with bestatin (Sigma, St. Louis, MO) to the final concentrations of 10 mM and 50 $\mu$g/ml, respectively. Bestatin was added to decrease the degradation of dipeptides in the extract (27). (Control experiments (not shown) showed that the addition of bestatin alone did not significantly inhibit the degradation of test proteins by the N-end rule pathway.) The following dipeptides and other amino acid derivatives were used: Arg-$\beta$-Ala, Ala-Lys, Lys-Ala, Trp-Ala (Sigma), and Tyr-His (Bachem Science, King of Prussia, PA). Stock samples of dipeptides were 0.5 M solutions in 10 mM K-Hepes, pH 7.5.

To follow the degradation of test proteins, reaction mixtures (0.1 ml) prepared as described above were incubated at 23°C. Samples (2.5 $\mu$l) were withdrawn in duplicate for each time point. One sample was examined by SDS-12% PAGE and autoradiography, using PhosphorImager (Molecular Dynamics, Sunnyvale, CA) (33). The other sample was used to determine the relative amount (%) of $^{35}$S soluble in 5% trichloroacetic acid (TCA) (30). This parameter was calculated as follows:

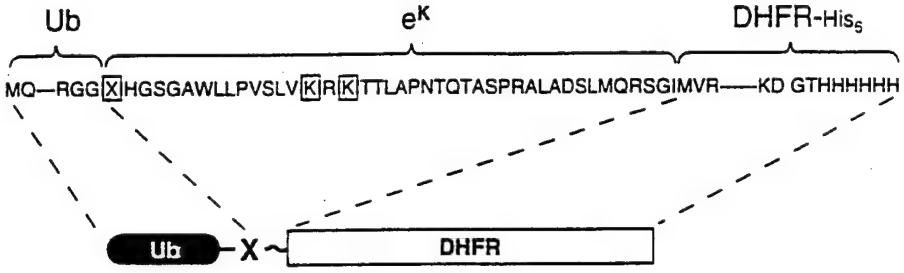



**FIG. 1.** The Ub-X-DHFR test proteins. These fusions, identical except for the variable residue X adjacent to the Ub moiety, were expressed and radiolabeled in *E. coli*, purified, and used as described under Materials and Methods. X was either Met, Arg, Leu, Phe, Cys, Asp, or Asn. The *E. coli* Lac repressor-derived 42-residue sequence, denoted as $e^K$ (extension (e) containing lysines (K)) was described previously (8, 30, 37). The alternative ubiquitylation sites Lys-15 and Lys-17 are boxed in the sequence of $e^K$ (8). DHFR represents DHFR-$His_6$, the mouse DHFR whose C-terminus was extended by 8 residues that included the $His_6$ tag.

$$\text{percent TCA-soluble } ^{35}\text{S} = \frac{X}{Y} \cdot \frac{a}{a-1} \cdot 100\%,$$

where $X$ is the amount of TCA-soluble $^{35}$S (cpm); $Y$ is the total amount of $^{35}$S (cpm) in the same sample; $a$ is the number of Met residues in a Ub-X-DHFR test protein ($a = 10$ for Ub-Met-DHFR; $a = 9$ for the other Ub-X-DHFRs). The $a - 1$ term above corrects for the presence of one Met residue in Ub.

The data presented in each of the experimental Figures were produced using the same extract on the same day. Unless stated otherwise, the data were consistent between at least two (usually three or more) independent experiments carried out with different preparations of the extract.

*In vitro apoptosis assays.* Apoptotic extracts from *Xenopus* eggs were prepared according to Newmeyer *et al.* (34). The difference between control and apoptotic egg extracts was in the time intervals between hormone injections used to produce the eggs. Frogs were injected with 75 units of pregnant mare serum gonadotropin (Calbiochem, San Diego, CA) 14–30 days (in contrast to 3–10 days for control extracts) before they were induced to lay eggs by injecting 800 units of human chorionic gonadotropin (Sigma). Eggs were collected, and the extract was prepared as described (32). The extracts were used within 1 h after preparation. For most assays, the extracts were arrested in interphase by the addition of 0.4 mM $CaCl_2$ and 0.1 mg/ml cycloheximide. Demembranated *Xenopus* sperm nuclei (32), at ~1000 per $\mu$l of extract, were also added at this time. Extracts treated as described were distributed into 0.1-ml samples and incubated at 23°C. Apoptotic changes in the extracts were monitored in two ways: by observing shrinkage and disintegration of the added sperm nuclei using phase contrast and fluorescent microscopy (34) and by measuring the DEVD-specific protease activity. The latter assay determined the total activity of caspases that recognize the tetrapeptide sequence DEVD (Asp-Glu-Val-Asp) (35, 36).

The DEVD-specific protease activity was measured by incubating 5 $\mu$l of extract with a quenched fluorescent substrate Ac-DEVD-AMC (50 $\mu$M) (Bachem Bioscience) in 0.1 ml of the extract buffer (15 mM $MgCl_2$, 1 mM DTT, 20 mM EGTA, 80 mM Na-$\beta$-glycerolphosphate, pH 7.3). The reaction was carried out for 10 min at 23°C, followed by freezing of a 20-$\mu$l sample in liquid $N_2$. Samples were thawed by diluting into 1 ml of phosphate-buffered saline, and their fluorescence (excitation at 380 nm and emission at 460 nm) was measured in the Hitachi F-4500 spectrofluorimeter (35, 36). AMC was used as a fluorescence standard.

## RESULTS

### *The N-End Rule Pathway in Xenopus Egg Extracts*

Among the cells of multicellular organisms, the N-end rule pathway has been identified and analyzed in rabbit reticulocytes (19) and L cells, a line of transformed, fibroblast-like mouse cells (26). To determine whether the N-end rule pathway is present in germ line cells such as *Xenopus* eggs, we used an approach similar to the earlier one with rabbit reticulocytes (19, 30).

A set of N-end rule substrates differing exclusively by their N-terminal residues was constructed, expressed in *E. coli*, metabolically labeled with [$^{35}$S]methionine, and purified by affinity chromatography (see Materials and Methods). A test protein of this set, Ub-X-$e^K$-DHFR-$His_6$, denoted below as Ub-X-DHFR, contained the following parts: the N-terminal Ub moiety; a variable residue X; $e^K$, a 42-residue, *E. coli* Lac repressor-derived sequence that contained the second (lysine) determinant of the N-degron (8, 37); and the $His_6$-tagged mouse dihydrofolate reductase moiety (Fig. 1). Ub fusions are not cleaved at the Ub-protein junction in *E. coli*, which lacks the Ub system (38). By contrast, in eukaryotes the Ub fusions, including the Ub-X-DHFRs, are rapidly cleaved by UBPs after the last residue of Ub, making it possible to produce, *in vivo* or *in vitro*, otherwise identical test proteins such as X-DHFRs that differ exclusively by their N-terminal residues (6, 7).

CSF-arrested *Xenopus* egg extracts were prepared from unfertilized frog eggs (32). Specific $^{35}$S-labeled Ub-X-DHFR test proteins (Fig. 1) were added to the egg extracts, and their metabolic fates were monitored either by SDS–PAGE and autoradiography or by measuring the amount of TCA-soluble $^{35}$S released during incubation at 23°C. The identity of residue X in the tested X-DHFRs encompassed at least one representative of each class of N-terminal destabilizing residues: Asn (tertiary destabilizing); Asp and Cys (secondary destabilizing), Arg (type 1 primary destabilizing), Phe and Leu (type 2 primary destabilizing), and Met (stabilizing) (see introduction) (7).

As expected, a Ub-X-DHFR test protein was deubiquitylated upon its addition to the extract. This was detected through an increase in the electrophoretic mobility of the major $^{35}$S-labeled species and the con-

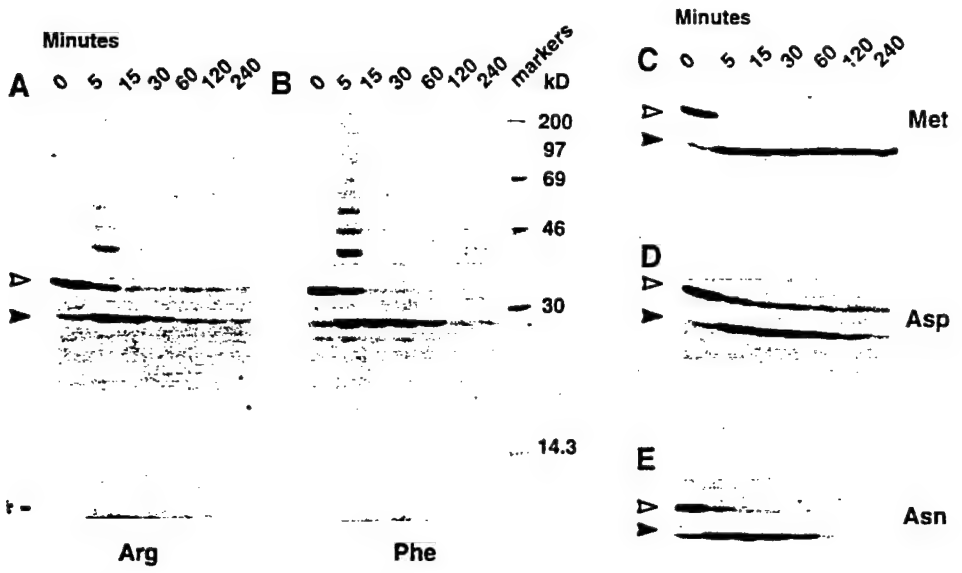



**FIG. 2.** Deubiquitylation and degradation of Ub-X-DHFR test proteins in *Xenopus* egg extract. (A) $^{35}$S-labeled, purified Ub-Arg-DHFR was added to a CSF-arrested *Xenopus* egg extract and incubated at 23°C for the indicated times. The samples were analyzed by SDS–PAGE, and the labeled species were detected by autoradiography (see Materials and Methods). (B) Same as A, but with Ub-Phe-DHFR. (C) Same as A, but with Ub-Met-DHFR (only the area around Met-DHFR is shown). (D) Same as C, but with Ub-Asp-DHFR. (E) Same as C, but with Ub-Asn-DHFR. Open arrowheads indicate the bands of the initial, 34-kDa Ub-X-DHFR. Closed arrowheads indicate the bands of deubiquitylated, 26-kDa X-DHFR. An asterisk indicates the band of free [$^{35}$S]Ub, which migrated close to the dye front in this electrophoretic system. The molecular masses (in kDa) of protein markers are indicated to the right of B.

comitant appearance of a labeled ~8 kDa species (Fig. 2 and data not shown). Deubiquitylation of Ub-X-DHFR in the extract was essentially complete in 5 min or less (Fig. 2). Whereas the relative intensity of the free Ub band remained nearly constant during the subsequent 4-h incubation (Figs. 2A and 2B, and data not shown), the relative intensity of the X-DHFR band was either constant or decreased, at a varying rate, depending on the identity of the N-terminal residue X. Based on their relative metabolic stabilities in the extract (Figs. 2 and 3A, and data not shown), the X-DHFR test proteins could be reproducibly ranked from the long-lived to the short-lived as follows: Met > Cys > Leu > Asp > Asn > Phe > Arg. We conclude that *Xenopus* egg extracts contain an N-end rule pathway whose rule book is similar, and may be identical, to those found in rabbit reticulocytes (19) and mouse L cells (26).

The decay curves of X-DHFRs that were determined by monitoring the release of TCA-soluble $^{35}$S and the resulting ranking of their N-terminal destabilizing residues were reproducible between experiments that utilized independent preparations of the egg extract (data not shown). Specifically, it took ~10 min to destroy 50% of Arg-DHFR to acid-soluble fragments. By 30 min, the degradation of Arg-DHFR was ~90% complete (Figs. 2A and 3A).

Approximately 20% of the added Met-DHFR, which bore a stabilizing N-terminal residue (7), was degraded during the first 10 min of incubation, and another 5 to 15% was degraded over the next 30 min. However, the rest of the Met-DHFR remained stable in the extract (Figs. 2C and 3A). A likely explanation of this result is that our preparation contained two types of Met-DHFR molecules: ~70% were undamaged and long-lived in the extract, as observed with Met-DHFR *in vivo,* in both *S. cerevisiae* and mouse L cells (Ref. 8; F. Lévy and A. V., unpubl. data), whereas the remainder comprised misfolded or otherwise damaged Met-DHFR molecules that were degraded by a proteolytic system distinct from the N-end rule pathway. By inference, the same background of degradation of a subpopulation of molecules would be expected for other X-DHFRs. It is unclear whether a subpopulation of misfolded or otherwise damaged X-DHFRs resulted from their overexpression in *E. coli,* or whether the damage occurred largely during purification of X-DHFRs. Independent evidence for this explanation of the background degradation of Met-DHFR is presented below.

Previous work has shown that N-terminal Cys is a stabilizing residue in *S. cerevisiae* (8), but is a secondary destabilizing residue in rabbit reticulocyte extracts and mouse L cells (19, 26). In *Xenopus* egg extracts, Cys was found to be a weakly destabilizing residue. Specifically, among the tested X-DHFRs, Cys-DHFR was the longest-lived test protein save for Met-DHFR, which bore a stabilizing N-terminal residue (Fig. 3A). Substrates that bear secondary or tertiary destabilizing residues require the tRNA-dependent activity of R-transferase for their degradation by the N-end rule

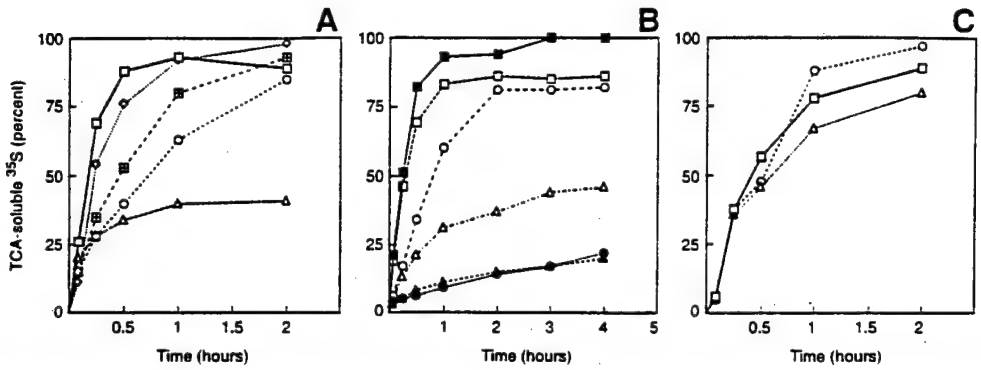



**FIG. 3.** Degradation of X-DHFR test proteins in egg extracts under different conditions. as determined by measuring TCA-soluble $^{35}$S. (A) $^{35}$S-labeled, purified Ub-Arg-DHFR (□), Ub-Phe-DHFR (◇), Ub-Met-DHFR (△), Ub-Asp-DHFR (○), and Ub-Asn-DHFR (⊞) were added to samples of the CSF-arrested *Xenopus* egg extract. After incubation at 23°C for the indicated times, the amounts of $^{35}$S soluble in 5% TCA were determined for each time point (see Materials and Methods). Since the free [$^{35}$S]Ub, produced through deubiquitylation of a Ub-X-DHFR, was stable during the incubation, 100% of TCA-soluble $^{35}$S represented complete degradation of the Ub-lacking [$^{35}$S]X-DHFR (see Materials and Methods for further details). (B) Pretreatment of egg extract with RNAase selectively inhibits degradation of N-end rule substrates bearing secondary destabilizing N-terminal residues. A CSF-arrested extract was pretreated with 5 units/ml of RNAase A-agarose (Sigma) for 1 h at 23°C, followed by the removal of RNAase-agarose by centrifugation (*closed symbol* curves). Another sample of the same extract was incubated for 1 h without RNAase (*open symbol* curves). $^{35}$S-labeled Ub-Cys-DHFR (*triangles*), Ub-Asp-DHFR (*circles*), or Ub-Arg-DHFR (*squares*) were added to both extracts, and the relative amounts of TCA-soluble $^{35}$S released during the incubation were determined as in A. (C) Different states of egg extract do not significantly alter the activity of the N-end rule pathway. $^{35}$S-labeled Ub-Phe-DHFR was added to the CSF-arrested egg extract (□), to the $CaCl_2$-treated extract that proceeded toward mitosis (△), and to the $CaCl_2$- and cycloheximide-treated, interphase-arrested extract (○). The degradation of Phe-DHFR was monitored as described in A.

pathway (see introduction). Previous work has shown that pretreatment of reticulocyte extract with RNAase A does not significantly affect the degradation of N-end rule substrates bearing primary destabilizing residues but abolishes the degradation of otherwise identical substrates bearing secondary or tertiary destabilizing N-terminal residues (19). In agreement with these findings, preincubation of *Xenopus* egg extract with RNAase A selectively inhibited the degradation of Cys-DHFR and Asp-DHFR, but had no effect on the degradation of Arg-DHFR (Fig. 3B). We conclude that both Cys and Asp are secondary destabilizing residues in the N-end rule pathway of *Xenopus* eggs. Taken together with the earlier findings about the N-end rule pathway in reticulocyte extract (19), the present findings are consistent with the conjecture that the set of secondary and tertiary destabilizing residues in the N-end rule of *Xenopus* eggs is identical to that of rabbit reticulocytes: two tertiary destabilizing residues (Asn and Gln) and three secondary ones (Asp, Glu, and Cys).

We also asked whether the rate of degradation of an N-end rule substrate depends on physiologically relevant changes in the state of an egg extract. Unfertilized *Xenopus* eggs are arrested at the metaphase of meiosis II through the action of cytostatic factor (39, 40). Fertilization results in a transient increase of the $Ca^{2+}$ concentration in the eggs that inactivates CSF, allowing cells to enter interphase and proceed to mitosis. This process can be mimicked *in vitro* by the addition of $Ca^{2+}$ (32). Such extracts, referred to as cycling extracts, undergo changes characteristic of eggs at mitosis, as observed through alterations in the morphology of *Xenopus* sperm nuclei added to the extract. The addition of cycloheximide and $CaCl_2$ to a CSF-arrested extract results not only in the escape from CSF arrest but also in a subsequent arrest at interphase, owing to the inhibition of protein synthesis, specifically the synthesis of cyclins (32). We examined the degradation of Phe-DHFR, bearing a type 2 primary destabilizing N-terminal residue, in the CSF-arrested, interphase-arrested, and cycling egg extracts, and found no significant differences in the rate of degradation of this N-end rule substrate in different extracts (Fig. 3C).

### *Dipeptides Bearing Destabilizing N-Terminal Residues Are Efficacious Inhibitors of the N-End Rule Pathway in Xenopus Egg Extracts*

Previous work has shown that the addition of amino acid derivatives such as dipeptides bearing destabilizing N-terminal residues to reticulocyte extracts or intact yeast cells results in a strong and selective inhibition of the N-end rule pathway in these settings (19, 27). However, experiments with intact mammalian cells in culture have shown dipeptides to be at most weak inhibitors of the N-end rule pathway in this setting[3]. The present work stemmed in part from the possibility of exploring the physiological consequences of perturbing a metazoan N-end rule pathway in the inhibitor-accessible setting of a cell-free system such as a *Xenopus* egg extract.

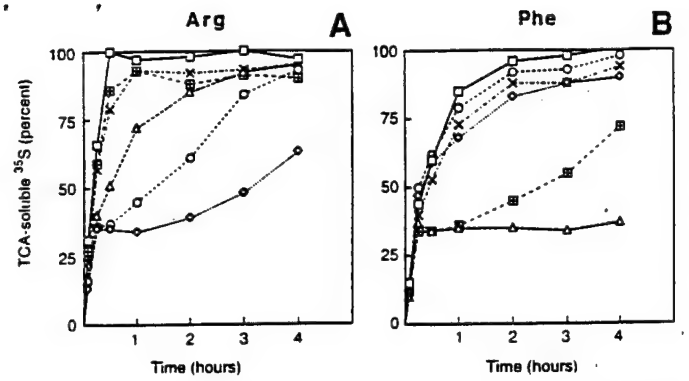



FIG. 4. Inhibition of the N-end rule pathway in egg extracts by dipeptides bearing destabilizing N-terminal residues. (A) $^{35}$S-labeled, purified Ub-Arg-DHFR was added to CSF-arrested *Xenopus* egg extracts in the absence (□) or the presence of the following dipeptides (10 mM), together with bestatin (50 μg/ml): Arg-β-Ala (◇), Lys-Ala (○), Tyr-His (△), Trp-Ala (⊞), and Ala-Lys (×). After incubation at 23°C for the indicated times, the amounts of $^{35}$S soluble in 5% TCA were determined for each time point. (B) Same as in A but with Ub-Phe-DHFR.

As shown in Figs. 4 and 5, Lys-Ala and especially Arg-β-Ala, both of which bear a type 1 primary destabilizing N-terminal residue, strongly inhibited the degradation of Arg-DHFR in the egg extract. This inhibition was selective, in that the same dipeptides had no significant effect on the degradation of Phe-DHFR, which bore a type 2 primary destabilizing N-terminal residue. Conversely, Tyr-His and Trp-Ala, which bear type 2 destabilizing N-terminal residues, strongly inhibited the degradation of Phe-DHFR, but had at most a small effect on the degradation of Arg-DHFR (Figs. 4 and 5). Ala-Lys, which bears a type 3 destabilizing N-terminal residue (19), did not inhibit the degradation of either Arg- or Phe-DHFRs (Fig. 4).

The dipeptides had little effect on the degradation of their cognate N-end rule substrates during the first 15 min of the incubation, but subsequent proteolysis was strongly inhibited (Fig. 4). This result is consistent with the finding that 20–30% of Met-DHFR, which bears a stabilizing N-terminal residue, was rapidly degraded, whereas the rest of the added Met-DHFR was stable (Fig. 3A). As described above, the latter finding could be accounted for if our preparation of Met-DHFR (and, by inference, of other X-DHFRs) comprised two distinct populations of molecules: undamaged ones (~70%) and hence, in the case of Met-DHFR, long-lived in the extract, and misfolded or otherwise damaged molecules that were recognized and degraded by a proteolytic system distinct from the N-end rule pathway. The failure of dipeptides to inhibit the initial burst of degradation of their cognate X-DHFRs (Fig. 4) is in agreement with this explanation.

*Apoptosis and the N-End Rule Pathway*

Several groups have described the use of *Xenopus* egg extracts to analyze the process of apoptosis (34, 36, 41–43). After several hours of incubation of the interphase-arrested, apoptotic egg extracts, the activity of DEVD-specific caspases (a subgroup in the ICE/CED-3 family of cysteine proteases) (43) rises sharply, followed by fragmentation of the added sperm nuclei (34). Both the induction of caspases and the fragmentation of the added nuclei in the egg extract can be suppressed by the addition of Bcl2, a protein that inhibits apoptosis *in vivo* (34, 36). Recent work has shown that apoptotic changes in egg extract are triggered, through unknown cytosolic factors, by the release of cytochrome *c* from the mitochondrial fraction of the extract (43). Apparently Bcl2 inhibits apoptosis by blocking the release of cytochrome *c* from mitochondria (43, 44).

We wished to determine whether the apoptosis-like events in egg extracts would be perturbed by dipeptide inhibitors of the N-end rule pathway. In preliminary experiments, we found that the times of apoptotic changes in egg extracts varied considerably between independent extract preparations. Specifically, the sharp rise of the DEVD-specific protease activity and the disintegration of the added sperm nuclei could occur as early as 2 h after the start of incubation or as

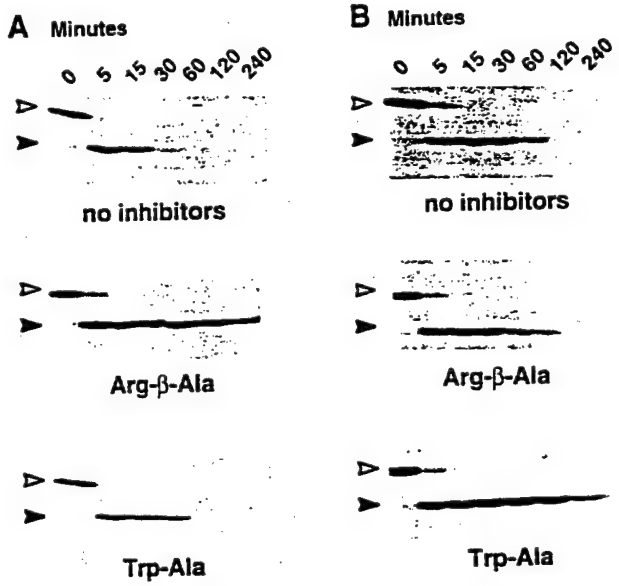



FIG. 5. Electrophoretic analysis of N-end rule substrates in egg extracts in the presence of dipeptides. (A) $^{35}$S-labeled, purified Ub-Arg-DHFR was added to a CSF-arrested *Xenopus* egg extract in the absence or the presence of dipeptides Arg-β-Ala or Trp-Ala (at 10 mM, together with bestatin at 50 μg/ml). After incubation at 23°C for the indicated times, the samples were analyzed by SDS–PAGE, and the labeled proteins were detected by autoradiography. Open arrowheads indicate the bands of the initial, 34-kDa Ub-Arg-DHFR. Closed arrowheads indicate the bands of deubiquitylated, 26-kDa Arg-DHFR. (B) Same as in A but with Ub-Phe-DHFR.

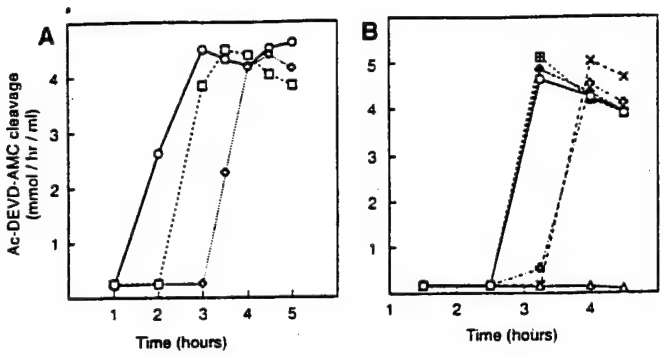



**FIG. 6.** Apoptotic changes in *Xenopus* egg extracts, assayed by measuring the DEVD-specific protease activity. (A) A preparation of the CSF-arrested egg extract was divided into three samples. The extracts were left untreated (□) or were activated by either 0.4 mM $CaCl_2$ (◇) or 0.4 mM $CaCl_2$ in the presence of 0.1 mg/ml cycloheximide (○). The samples were incubated at room temperature for up to 5 h. The DEVD-specific protease activity in the extracts was determined using the fluorogenic substrate Ac-DEVD-AMC. (B) Samples of interphase-arrested egg extract were incubated either in the absence (○) or in the presence of the following reagents: Lys-Ala at 20 mM (⊞); Ala-Lys at 20 mM (◆); Trp-Ala at 20 mM (×); Trp-Ala plus Lys-Ala, each at 10 mM (⊕); Ac-DEVD-CHO (an inhibitor of DEVD proteases, at 10 mM) (△). The dipeptides were added together with bestatin (50 μg/ml). The DEVD-specific protease activity was measured as in A.

late as 5 h (data not shown). This variability, reported by others as well (43), is likely to result in part from the variability of the timing of upstream events that trigger the beginning of apoptosis in the *Xenopus* egg extract. In addition, the timing of apoptotic changes depended on the state of the extracts: when a single preparation was split into three samples, and was used as CSF-arrested, cycling, and interphase-arrested extracts, the increase of the DEVD-specific protease activity was observed first in the interphase-arrested extract, then in the CSF-arrested one, and later in the cycling extract (Fig. 6A). This pattern was reproducible between different preparations of the extract, even though the absolute timing of the rise of the DEVD-specific protease activity varied significantly between preparations (data not shown).

We asked whether selective inhibition of the N-end rule pathway would affect the nature or kinetics of the apoptotic changes. Dipeptides that had been shown to be efficacious inhibitors of the N-end rule pathway (Figs. 2 and 3A) were added to the interphase-arrested and cycling extracts at the beginning of incubation. We found that Lys-Ala and Arg-β-Ala, containing type 1 primary destabilizing N-terminal residues, and Ala-Lys, containing a type 3 destabilizing residue, did not alter the timing of DEVD-specific protease induction (Fig. 6B). In contrast, Trp-Ala, which bears a type 2 destabilizing N-terminal residue, delayed both the rise of the DEVD-specific protease activity and the disintegration of sperm nuclei by ~1 h in 8 out of 12 independent experiments (Fig. 6B and data not shown). The ~1 h delay of apoptotic changes by Trp-Ala was significant but weak in comparison to the effect of Ac-DEVD-CHO, an inhibitor of DEVD-specific proteases, which delayed apoptosis by at least 2 h (data not shown). Interestingly, the proteasome inhibitor 5-iodo-4-hydroxyl-3-nitrophenylacetyl-leucinyl-leucinyl-leucine-vinyl sulfone (NIP-$L_3$VS) (45), which inhibited the degradation of Phe-DHFR in egg extracts, also delayed apoptosis in the CSF-arrested extract by ~1 h (data not shown), similarly to the effect of Trp-Ala (Fig. 6B).

## DISCUSSION

We report the following results:

(1) A whole-cell extract of *X. laevis* eggs contains the N-end rule pathway, as determined by monitoring the metabolic fates of purified, $^{35}$S-labeled Ub-X-DHFR test proteins added to the extract. Among the cells of multicellular organisms, the N-end rule pathway has been identified and characterized in rabbit reticulocytes (18) and L cells, a mouse cell line (19, 26), but not, until now, in germ line cells.

(2) The partial N-end rule defined in egg extracts was found to be identical to the corresponding subsets of the N-end rules in rabbit reticulocytes and mouse fibroblasts. In particular, we showed that in egg extracts the destabilizing activity of N-terminal Asn, Asp, and Cys requires the presence of RNA (presumably tRNA), strongly suggesting that these are secondary (Asp and Cys) and tertiary (Asn) residues in the egg's N-end rule pathway, as they have been shown to be in other metazoan cells (see introduction for the terminology) (7).

(3) The activity of the N-end rule pathway did not change significantly upon shifts of the CSF-arrested egg extract to the cycling or the interphase-arrested state.

(4) Dipeptides bearing destabilizing N-terminal residues selectively and efficiently inhibited the degradation of N-end rule substrates in egg extract. In particular, dipeptides bearing the basic (type 1) primary destabilizing N-terminal residues inhibited the degradation of N-end rule substrates bearing type 1 but not type 2 N-terminal residues. A converse inhibition pattern was observed with dipeptides bearing bulky hydrophobic (type 2) N-terminal residues.

(5) The onset of apoptosis-like changes in egg extract (fragmentation of the added sperm nuclei and abrupt increase of the DEVD-specific protease activity (34, 43)) was found to be delayed by ~1 h in the presence of dipeptide Trp-Ala, which bears a type 2 destabilizing N-terminal residue, but not in the presence of dipeptides bearing type 1 destabilizing N-terminal residues.

The N-end rule pathway has been identified in all organisms examined, from mammals to fungi and bac-

teria, but the understanding of its functions remains incomplete (6, 7). The first clear example of a physiological function of the N-end rule pathway is its recently identified role in controlling the import of peptides in *S. cerevisiae* through the degradation of the transcriptional repressor Cup9p that down-regulates a peptide transporter (25, 46). It has also been reported that dipeptides bearing destabilizing N-terminal residues specifically inhibit differentiation of rat pheochromocytoma PC12 cells (47), the neurite outgrowth in amphibian neuroepithelial cells (48), and limb regeneration in the newt (49), suggesting a role for the N-end rule pathway in these biological processes. Unfortunately, none of these studies (47–49) included control experiments to verify that the added dipeptides actually inhibited degradation of a reporter N-end rule substrate in the target cells. Thus, it remains to be determined whether the observed biological effects of dipeptides (47–49) were caused by specific inhibition of the N-end rule pathway in the target cells or resulted from other effects of the added dipeptides.

Apart from the interest in determining whether the N-end rule pathway is present in germ line cells such as *Xenopus* eggs, one reason for initiating this work was the opportunity to test, through selective inhibition of the N-end rule pathway in egg extracts, whether this pathway plays a role in apoptosis. Several lines of evidence suggest a role for the Ub/proteasome system, of which the N-end rule pathway is a part, in regulating apoptosis (50–55). In particular, cell-penetrating proteasome inhibitors were reported to partially protect nonproliferating cells such as thymocytes and growth factor-deprived sympathetic neurons from apoptosis (52, 53). In contrast, the same proteasome inhibitors were reported to induce apoptosis in mitotically active cells (51, 54). An apoptosis-like process, induced by the stress of starvation and mediated by proteolysis, was also identified in *E. coli* (56). A key regulator of *E. coli* apoptosis is a short-lived protein MazE, which is degraded by ClpAP, a proteasome-like ATP-dependent protease that targets N-end rule substrates in *E. coli* (38). These and other considerations, including the possibility that cleavages by caspases can produce N-end rule substrates, led to the suggestion that the N-end rule pathway may function in the control of apoptosis (7).

Specific dipeptides, which have been shown to act as selective inhibitors of the N-end rule pathway in reticulocyte extracts (19, 27) and intact *S. cerevisiae* cells (28), were found to be largely ineffective with several lines of mammalian cells[3]. *Xenopus* egg extracts, in contrast to reticulocyte extracts, have been shown to undergo apoptosis-like changes (34, 36, 43), hence the choice of these extracts for the present study. We found that dipeptides bearing either type 1 (Arg, Lys) or type 3 (Ala) destabilizing N-terminal residues had no effect on the apoptotic changes in the extracts. In contrast, Trp-Ala, which bears a type 2 destabilizing N-terminal residue, inhibited apoptosis in a number of independent experiments with different preparations of egg extracts; in a minority of tests the effect was not observed. Tyr-His, a dipeptide bearing another type 2 destabilizing N-terminal residue, strongly inhibited the degradation of an N-end rule substrate such as Phe-DHFR in the extract, but was significantly less efficacious than Trp-Ala in its effect on the timing of apoptotic changes in the same extract. Taken together, the results of experiments with dipeptides as inhibitors of apoptosis indicate that some inhibitors of the N-end rule pathway, notably Trp-Ala, can cause a significant delay of apoptosis in the extract. A definitive test and further analysis of the thus suggested function of the N-end rule pathway in apoptosis will require mutants that eliminate this pathway in a multicellular organism without perturbing the rest of the Ub system. Construction of a mouse mutant that lacks the entire N-end rule pathway (through a deletion of the *Ubr1* gene that encodes N-recognin, the main recognition component of this proteolytic system) is under way[4].

In summary, we established the existence of the N-end rule pathway in germ line cells such as *Xenopus* eggs, showed that dipeptides are efficacious inhibitors of the N-end rule pathway in egg extracts, and examined the effects of dipeptides on apoptotic changes in these extracts.

## ACKNOWLEDGMENTS

We are grateful to W. Dunphy for helpful discussions and the access to his laboratory's *Xenopus* facility, and to M. Bogyo and H. Ploegh for a gift of the NIP-$L_3$VS proteasome inhibitor. We thank E. Smirnova and members of the Varshavsky lab, especially L. Peck, G. Turner, and A. Webster, for their suggestions and for the comments on the manuscript. D.P. is a Fellow of the Leukemia Society of America. This work was supported by grants to A.V. from the National Institutes of Health (DK39520 and GM31530).

## REFERENCES

1. Haas, A. J., and Siepman, T. J. (1997) *FASEB J.* **11,** 1257–1268.
2. Varshavsky, A. (1997) *Trends Biochem. Sci.* **22,** 383–387.
3. Hochstrasser, M. (1996) *Annu. Rev. Genet.* **30,** 405–439.
4. King, R. W., Deshaies, R. J., Peters, J. M., and Kirschner, M. W. (1996) *Science* **274,** 1652–1659.
5. Varshavsky, A. (1991) *Cell* **64,** 13–15.
6. Bachmair, A., Finley, D., and Varshavsky, A. (1986) *Science* **234,** 179–186.
7. Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 12142–12149.
8. Bachmair, A., and Varshavsky, A. (1989) *Cell* **56,** 1019–1032.
9. Johnson, E. S., Gonda, D. K., and Varshavsky, A. (1990) *Nature* **346,** 287–291.

[4] Y. T. Kwon and A. Varshavsky, unpublished data.

10. Hill, C. P., Johnston, N. L., and Cohen, R. E. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 4136–4140.
11. Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K., and Varshavsky, A. (1989) *Science* **243,** 1576–1583.
12. Pickart, C. M. (1997) *FASEB J.* **11,** 1055–1066.
13. Wilkinson, K. D. (1997) *FASEB J.* **11,** 1245–1256.
14. Hershko, A., and Ciechanover, A. (1992) *Annu. Rev. Biochem.* **61,** 761–807.
15. Finley, D., and Chau, V. (1991) *Annu. Rev. Cell Biol.* **7,** 25–69.
16. Coux, O., Tanaka, K., and Goldberg, A. L. (1996) *Annu. Rev. Biochem.* **65,** 801–817.
17. Hilt, W., and Wolf, D. H. (1996) *Trends. Biochem. Sci.* **21,** 96–102.
18. Lupas, A., Flanagan, J. M., Tamura, T., and Baumeister, W. (1997) *Trends Biochem. Sci.* **22,** 399–404.
19. Gonda, D. K., Bachmair, A., Wünning, I., Tobias, J. W., Lane, W. S., and Varshavsky, A. (1989) *J. Biol. Chem.* **264,** 16700–16712.
20. Balzi, E., Choder, M., Chen, W., Varshavsky, A., and Goffeau, A. (1990) *J. Biol. Chem.* **265,** 7464–7471.
21. Baker, R. T., and Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 12065–12074.
22. Bartel, B., Wünning, I., and Varshavsky, A. (1990) *EMBO J.* **9,** 3179–3189.
23. deGroot, R. J., Rümenapf, T., Kuhn, R. J., and Strauss, J. H. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 8967–8971.
24. Madura, K., and Varshavsky, A. (1994) *Science* **265,** 1454–1458.
25. Byrd, C., Turner, G. C., and Varshavsky, A. (1998) *EMBO J.* **17,** 269–277.
26. Lévy, F., Johnsson, N., Rümenapf, T., and Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 4907–4912.
27. Reiss, Y., Kaim, D., and Hershko, A. (1988) *J. Biol. Chem.* **263,** 2693–269.
28. Baker, R. T., and Varshavsky, A. (1991) *Proc. Natl. Acad. Sci. USA* **87,** 2374–2378.
29. Dunphy, W. G. (1994) *Trends Cell Biology* **4,** 202–207.
30. Johnston, J. A., Johnson, E. S., Waller, P. R. H., and Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 8172–8178.
31. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. Eds. (1996) Current Protocols in Molecular Biology, Wiley–Interscience, New York.
32. Murray, A. W. (1991) *Methods Cell. Biol.* **36,** 581–605.
33. Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N. A., Copeland, N. J., and Varshavsky, A. (1996) *J. Biol. Chem.* **271,** 28521–28532.
34. Newmeyer, D. D., Farschon, D. M., and Reed, J. C. (1994) *Cell* **79,** 353–364.
35. Nicholson, D. W., Ali, A., Thornberry, N. A., Vaillancourt, J. P., Ding, C. K., Gallant, M., Gareau, Y., Griffin, P. R., Labelle, M., Lazebnik, Y. A., Munday, N. A., Raju, S. M., Smulson, M. E., Yamin, T.-T., Yu, V. L., and Miller, D. K. (1995) *Nature* **376,** 37–43.
36. Cosulich, S. C., Green, S., and Clarke, P. R. (1996) *Curr. Biol.* **6,** 353–364.
37. Johnson, E. S., Ma, P. C. M., Ota, I. M., and Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 17442–17456.
38. Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. (1991) *Science* **254,** 1374–1377.
39. Sagata, N. (1996) *Trends Cell Biol.* **6,** 22–28.
40. Furuno, N., Ogawa, Y., Iwashita, J., Nakajo, N., and Sagata, N. (1997) *EMBO J.* **16,** 3860–3865.
41. Evans, E. K., Lu. W., Strum, S. L., Mayer, B. J., and Kornbluth, S. (1997) *EMBO J.* **16,** 230–241.
42. Kluck, R. M., Bossy-Wetzel, E., Green, D. R., and Newmeyer, D. D. (1997) *Science* **275,** 1132–1136.
43. Kluck, R. M., Martin, S. J., Hoffman, B. M., Zhou, J. S., Green, D. R., and Newmeyer, D. D. (1997) *EMBO J.* **16,** 4639–4649.
44. Villa, P., Kaufmann, S. H., and Earnshaw, W. C. (1997) *Trends Biochem. Sci.* **22,** 388–393.
45. Bogyo, M., McMaster, J. S., Gaczynska, M., Tortorella, D., Goldberg, A. L., and Ploegh, H. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 6629–6634.
46. Alagramam, K., Naider, F., and Becker, J. M. (1995) *Mol. Microbiol.* **15,** 225–234.
47. Hondermarck, H., Sy, J., Bradshaw, R. A., and Arfin, S. M. (1992) *Biochem. Biophys. Res. Commun.* **30,** 280–288.
48. Maufroid, J. P., Bradshaw, R. A., Boilly, B., and Hondermarck, H. (1996) *Int. J. Dev. Biol.* **40,** 609–611.
49. Taban, C. H., Hondermarck, H., Bradshaw, R. A., and Boilly, B. (1996) *Experentia* **52,** 865–870.
50. Delic, J., Morange, M., and Magdelenat, H. (1993) *Mol. Cell. Biol.* **13,** 4875–4883.
51. Imajoh-Ohmi, S., Kawaguchi, T., Sugiyama, S., Tanaka, K., Omura, S., and Kikuchi, H. (1995) *Biochem. Biophys. Res. Commun.* **217,** 1070–1077.
52. Grimm, L. M., Goldberg, A. L., Poirier, G. G., Schwartz, L. M., and Osborne, B. A. (1996) *EMBO J.* **15,** 3835–3844.
53. Sadoul, R., Fernandez, P.-A., Quiquerez, A.-L., Martinou, I., Maki, M., Schröter, M., Becherer, J. D., Irmler, M., Tschopp, J., and Martinou, J.-C. (1996) *EMBO J.* **15,** 3845–3852.
54. Drexler, H. C. A. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 855–860.
55. Lopes, U. G., Erhardt, P., Yao, R., and Cooper, G. M. *J. Biol. Chem.* **272,** 12893–12896.
56. Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 6059–6063.

The EMBO Journal Vol.17 No.1 pp.269–277, 1998

# The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor

**Christopher Byrd, Glenn C.Turner and Alexander Varshavsky[1]**

Division of Biology, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

[1]Corresponding author
e-mail: avarsh@cco.caltech.edu

**Ubiquitin-dependent proteolytic systems underlie many processes, including the cell cycle, cell differentiation and responses to stress. One such system is the N-end rule pathway, which targets proteins bearing destabilizing N-terminal residues. Here we report that Ubr1p, the main recognition component of this pathway, regulates peptide import in the yeast *Saccharomyces cerevisiae* through degradation of Cup9p, a 35 kDa homeodomain protein. Cup9p was identified using a screen for mutants that bypass the previously observed requirement for Ubr1p in peptide import. We show that Cup9p is a short-lived protein ($t_{1/2}$ ~5 min) whose degradation requires Ubr1p. Cup9p acts as a repressor of *PTR2*, a gene encoding the transmembrane peptide transporter. In contrast to engineered N-end rule substrates, which are recognized by Ubr1p through their destabilizing N-terminal residues, Cup9p is targeted by Ubr1p through an internal degradation signal. The Ubr1p–Cup9p–Ptr2p circuit is the first example of a physiological process controlled by the N-end rule pathway. An earlier study identified Cup9p as a protein required for an aspect of resistance to copper toxicity in *S.cerevisiae*. Thus, one physiological substrate of the N-end rule pathway functions as both a repressor of peptide import and a regulator of copper homeostasis.**

*Keywords*: *CUP9*/N-end rule/peptide import/proteolysis/*PTR2*/*UBR1*

## Introduction

Many regulatory proteins are short-lived *in vivo* (Schwob *et al.*, 1994; King *et al.*, 1996; Varshavsky, 1996). This metabolic instability makes possible rapid adjustment of the protein's concentration (or subunit composition) through changes in the rates of its synthesis or degradation. Protein degradation plays a role in a multitude of processes, including cell growth, division, differentiation and responses to stress. In eukaryotes, a large fraction of intracellular proteolysis is mediated by the ubiquitin system. Ubiquitin (Ub) is a 76-residue protein whose covalent conjugation to other proteins marks them for processive degradation by the 26S proteasome—an ATP-dependent, multisubunit protease (Jentsch and Schlenker, 1995; Hilt and Wolf, 1996; Hochstrasser, 1996; Rubin *et al.*, 1997).

Features of proteins that confer metabolic instability are called degradation signals (degrons). One of the degradation signals recognized by the ubiquitin system is called the N-degron. It comprises two essential determinants: a destabilizing N-terminal residue and an internal lysine of a substrate (Bachmair *et al.*, 1986; Varshavsky, 1996). The Lys residue is the site of formation of a substrate-linked multiubiquitin chain (Bachmair and Varshavsky, 1989; Chau *et al.*, 1989). A set of N-degrons bearing different N-terminal residues that are destabilizing in a given cell type yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct versions of the N-end rule operate in all organisms examined, from mammals to fungi and bacteria (Varshavsky, 1996).

The N-end rule pathway is organized hierarchically. In eukaryotes such as *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase (Nt-amidase), into the secondary destabilizing residues Asp and Glu (Baker and Varshavsky, 1995). Secondary residues, in turn, function through their conjugation to Arg by the *ATE1*-encoded Arg-tRNA–protein transferase (R-transferase) (Balzi *et al.*, 1990). Arg is one of several primary destabilizing N-terminal residues which are bound directly by N-recognin, a 225 kDa E3 protein encoded by the *UBR1* gene (Bartel *et al.*, 1990). Ubr1p, together with the associated ubiquitin-conjugating (E2) enzyme Ubc2p, mediates the formation of a substrate-linked multiubiquitin chain (Varshavsky, 1996).

The N-end rule pathway was first encountered in experiments that explored, in *S.cerevisiae*, the metabolic fate of a fusion between Ub and a reporter such as *Escherichia coli* β-galactosidase (β-gal) (Bachmair *et al.*, 1986). While such engineered N-end rule substrates have been extensively characterized (Varshavsky, 1996), little is known about their physiological counterparts. The few identified so far include the *GPA1*-encoded Gα subunit of the *S.cerevisiae* heterotrimeric G protein, which mediates the pheromone response in this fungus, and RNA polymerases of alphaviruses whose hosts include mammalian and insect cells (de Groot *et al.*, 1991; Madura and Varshavsky, 1994). Physiological functions of the instability of these proteins remain to be understood (Varshavsky, 1996). Inactivation of the N-end rule pathway in *S.cerevisiae*—through deletion of the *UBR1* gene—results in cells which grow slightly slower than their wild-type counterparts, and are impaired in sporulation (increased frequency of asci containing fewer than four spores), but otherwise appear to be normal (Bartel *et al.*, 1990).

Recently, Becker and colleagues (Alagramam *et al.*, 1995) have reported that *ubr1*Δ cells are deficient in the

A

| Strain | Genotype | Peptide import | Strain | Genotype | Peptide import |
|---|---|---|---|---|---|
| JD53 | wt | + | MHY599 | *pas2Δ* | + |
| JD83-1A | *ubr1Δ* | – | CBY8 | *ubc2Δ, ubc1Δ* | + |
| KM207-1 | *ate1Δ* | + | CBY11 | *ubc2Δ, cdc34-1*[1] | + |
| SGY6 | *nta1Δ* | + | EJY105/6 | *ubc2Δ, ubc4Δ* | – |
| CBY20 | *ptr2Δ* | + | CBY4 | *ubc2Δ, ubc5Δ* | + |
| EJ2 | *ubc1Δ* | + | CBY5 | *ubc2Δ, ubc6Δ* | + |
| BBY67 | *ubc2Δ* | +/– | CBY6 | *ubc2Δ, ubc7Δ* | + |
| JD34 | *ubc2Δ* | +/– | CBY7 | *ubc2Δ, ubc8Δ* | + |
| EJY102 | *ubc2Δ* | +/– | CBY9 | *ubc2Δ, ubc9-1*[1] | + |
| RJD549 | *cdc34-1*[1] | + | CBY10 | *ubc2Δ, pas2Δ* | + |
| RJD795 | *cdc34-2*[1] | + | EJ10 | *ubc4Δ, ubc5Δ* | + |
| EJ3 | *ubc4Δ* | + | MHY503 | *ubc4Δ, ubc6Δ* | + |
| EJ1 | *ubc5Δ* | + | MHY554 | *ubc4Δ, ubc7Δ* | + |
| MHY495 | *ubc6Δ* | + | MHY552 | *ubc6Δ, ubc7Δ* | + |
| MHY507 | *ubc7Δ* | + | MHY550 | *ubc4Δ, 5Δ, 6Δ* | + |
| MHY601 | *ubc8Δ* | + | MHY557 | *ubc4Δ, 6Δ, 7Δ* | + |
| YWO55 | *ubc9-1*[1] | + | MHY570 | *ubc4Δ, 5Δ, 6Δ, 7Δ* | + |

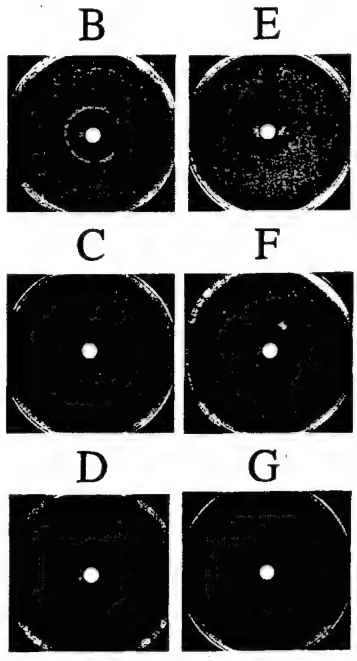



**Fig. 1.** The import of peptides is decreased in the absence of Ubc2p and virtually abolished in the absence of Ubc2p and Ubc4p. *Saccharomyces cerevisiae* mutants deficient in one or more ubiquitin-conjugating (E2) enzymes or other components of the ubiquitin system were tested for their ability to import peptides, using the halo assay and a toxic dipeptide L-leucyl-L-ethionine (Leu-Eth) (see Materials and methods). '+', '±' and '-' denote, respectively, the apparently wild-type, significantly reduced, and undetectable levels of peptide import. A superscript '1' refers to strains that carried a *ts* allele of an essential E2 enzyme, either Cdc34p (Ubc3p) or Ubc9p; the 30°C temperature of the test was semi-permissive for these strains. (**A**) Summary of the results. (B–F) Examples of the actual halo assays, with wild-type (**B**), *ptr2Δ* (**C**), *ubr1Δ* (**D**), *ubc2Δ* (**E**), *ubc4Δ* (**F**) and *ubc2Δ ubc4Δ* (**G**) strains of *S.cerevisiae*. Elimination of some E2 enzymes in the *ubc2Δ* background restored halo formation, presumably because such strains were growth-impaired in a way that made them hypersensitive to the toxicity of ethionine. By contrast, although *ubc2Δ ubc4Δ* cells were also growth-impaired, they were import-defective and therefore grew in the immediate vicinity of the filter.

import of di- and tripeptides, suggesting that this process, which is universal among living cells, requires the N-end rule pathway. In the present work, we identified the underlying regulatory mechanism and discovered a new physiological substrate of the N-end rule pathway, the homeodomain protein Cup9p. This short-lived protein is targeted for degradation by Ubr1p, and acts as a transcriptional repressor of *PTR2*, a gene that encodes a transmembrane peptide transporter. The Ubr1p–Cup9p–Ptr2p circuit is the first example of a physiological process controlled by the N-end rule pathway.

## Results

### *The involvement of Ubc2p and Ubc4p E2 enzymes in the control of peptide import*

We began by asking whether components of the N-end rule pathway other than Ubr1p were also necessary for the import of peptides. Previous work has shown that Ubc2p, one of 13 ubiquitin-conjugating (E2) enzymes of *S.cerevisiae*, is required for the degradation of engineered N-end rule substrates, and is physically associated with the E3 protein Ubr1p (N-recognin) (Jentsch, 1992; Madura *et al.*, 1993).

To test for the ability of *S.cerevisiae* to import peptides, we used a halo assay, in which a filter soaked in the toxic dipeptide L-leucyl-L-ethionine (Leu-Eth) is placed on a plate, inhibiting the growth of import-competent cells near the filter. By this test, the elimination of *UBC2* impaired, but did not abolish, the import of peptides (Figure 1A, B and E). To determine which of the other E2 enzymes, if any, were required for the residual peptide import observed in *ubc2Δ* cells, a number of single and multiple mutants in *UBC* genes were examined (Figure 1). We found that the elimination of both *UBC2* and *UBC4* virtually abolished the import of peptides (Figure 1A, B and E–G). Elimination of Ubc4p, one of the more abundant E2 enzymes (Bachmair *et al.*, 1986; Jentsch, 1992), had previously been noticed to decrease slightly the activity of the N-end rule pathway (Bartel, 1990). Cells lacking either Nta1p or Ate1p—the 'upstream' components of this pathway—were also examined and found unimpaired in the import of peptides, in contrast to *ubr1Δ* and *ubc2Δ ubc4Δ* cells (Figure 1A).

### *Identification of Cup9p as a negative regulator of peptide import*

Previous work (Alagramam *et al.*, 1995) has shown that deletion of *UBR1* greatly reduces the level of *PTR2* mRNA, which encodes the transmembrane peptide transporter. This result, and our observation that peptide import requires the presence of at least one of two specific ubiquitin-conjugating enzymes (Figure 1), suggested a model in which expression of the Ptr2p transporter is regulated by a short-lived repressor that is degraded by the N-end rule pathway. In cells lacking *UBR1*, the repressor would be expected to accumulate, thereby blocking peptide import. One prediction of this model is that inactivation of this repressor would bypass the requirement for Ubr1p in peptide import. A screen for such 'bypass' mutations (see Materials and methods) yielded 199 recessive isolates, of which 101 defined one complementation group, termed

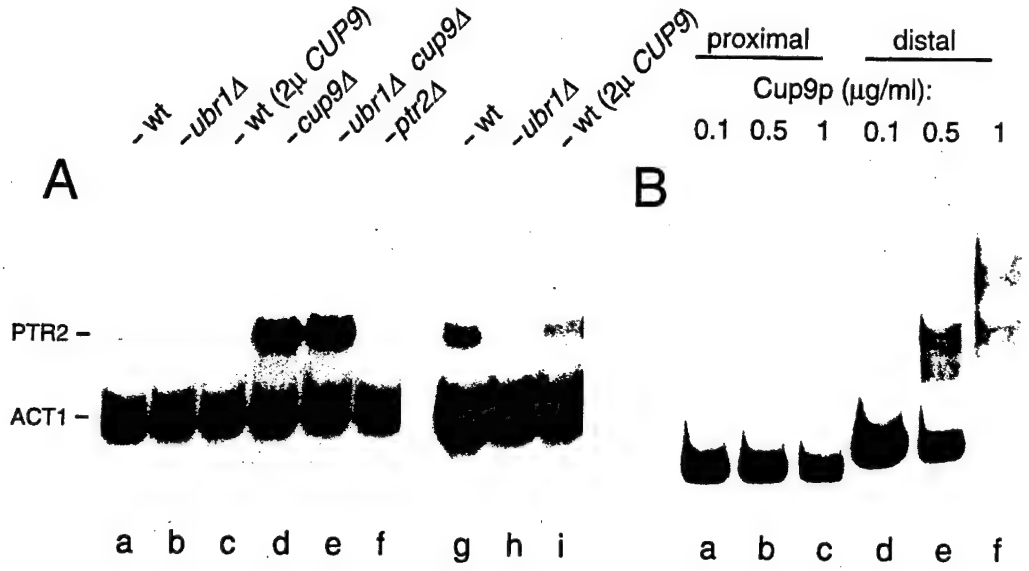



**Fig. 2.** Cup9p is a repressor of the *PTR2* gene. (**A**) Expression of the *S.cerevisiae PTR2* gene in different genetic backgrounds. Equal amounts of total RNA isolated from different strains were analyzed by Northern hybridization (see Materials and methods), using the *PTR2* (peptide transporter) and *ACT1* (actin) genes as $^{32}$P-labeled probes. Lane a, JD52 (*CUP9 UBR1*) (wild-type) transformed with pCB201 (empty vector). Lane b, JD55 (*CUP9 ubr1Δ*) transformed with pCB201. Lane c, JD52 (*CUP9 UBR1*) transformed with the high-copy plasmid pCB209 that expressed *CUP9* from its natural promoter. Lane d, CBY18 (*cup9Δ UBR1*). Lane e, CBY16 (*cup9Δ ubr1Δ*). Lane f, CBY21 (*CUP9 UBR1 ptr2Δ*). Lanes g–i, same as lanes a–c, but a longer autoradiographic exposure to highlight the tight repression of *PTR2* by Cup9p in *ubr1Δ* cells (b, h) and the difference between levels of *PTR2* mRNA in the wild-type cells (a, g) and their counterparts that overexpressed Cup9p (c, i). (**B**) Cup9p specifically binds to a site in the $P_{PTR2}$ promoter. A gel shift assay with Cup9-$H_6$, poly-dI·dC and $^{32}$P-labeled DNA fragments of the *PTR2* promoter hybridization (see Materials and methods). Lanes a–c, with a fragment (–1 to –447) proximal to the inferred start codon of the *PTR2* ORF. Lanes d–f, same as lanes a–c, but with a more distal DNA fragment (–448 to –897). Concentrations of Cup9-$H_6$ (in μg/ml) are indicated above the lanes.

*sub1* (suppressor of a block to peptide import in *ubr1Δ*). In agreement with the model's prediction, *sub1* mutants acquired the ability to express *PTR2* in the absence of *UBR1* (Figure 2A, lane b versus lane e). The *sub1* locus was cloned by complementation (see Materials and methods), and was found to be the *CUP9* gene.

*CUP9* was originally identified by Knight *et al.* (1994) as a gene whose disruption impairs the copper resistance of *S.cerevisiae* growing on lactate, a non-fermentable carbon source. Under these conditions, Cup9p plays a major (but mechanistically obscure) role in copper homeostasis (Knight *et al.*, 1994). *CUP9* encodes a 35 kDa protein that contains a homeodomain, an ~60-residue helix–turn–helix DNA-binding motif present in many eukaryotic regulatory proteins (Wolberger, 1996). Outside the homeodomain region, the sequence of Cup9p is not similar to sequences in databases.

To verify that *CUP9* and *SUB1* were the same gene, complementation tests were carried out. Two independently derived *ubr1Δ cup9::LEU2* strains (CBY16 and CBY17) were crossed to *ubr1Δ sub1-1* (CBY15), and the resulting diploids (CBY23 and CBY24, respectively) were tested for their ability to import dipeptides (see Materials and methods). As would be expected of allelic loci, *cup9::LEU2* and *sub1-1* failed to complement one another: both diploid strains remained import-competent (data not shown). In another test, CBY23 and CBY24 were sporulated, and the segregants were analyzed for the presence of the *LEU2* gene (integrated at the *CUP9* locus) and for the ability to import peptides. Among the eight tetrads tested, *LEU2* was present in two of the four segregants, whereas all four segregants were invariably import-competent, a pattern expected if *CUP9* and *SUB1* were one and the same gene.

To examine the regulation of peptide import by Cup9p and Ubr1p, congenic *S.cerevisiae* strains that lacked, expressed or overexpressed Cup9p and/or Ubr1p were constructed and assayed for peptide import by growth on selective media. Suspensions of cells (auxotrophic for lysine) were serially diluted and plated on either rich media, minimal media lacking lysine and containing Lys-Ala dipeptide (selecting for peptide import), or minimal media containing the toxic dipeptide Leu-Eth (selecting against peptide import). All strains grew at comparable rates on rich media (Figure 3A). On minimal media that supplied the essential lysine as the Lys-Ala dipeptide, the *CUP9 UBR1* (wild-type), *cup9Δ UBR1* and *cup9Δ ubr1Δ* strains grew at comparable rates (Figure 3B), whereas the *CUP9 ubr1Δ* strain failed to grow (Figure 3B), in agreement with the observation that *UBR1* is required for peptide import (Figure 1). Opposite growth patterns were observed on media containing toxic dipeptide (Figure 3C). Comparison of the data in Figure 3 with *PTR2* mRNA levels (Figure 2A) suggested that *CUP9* exerts its effect on the import of peptides by repressing transcription of *PTR2*. For example, wild-type (*CUP9 UBR1*) cells overexpressing Cup9p exhibited reduced levels of *PTR2* mRNA (Figure 2A) and decreased sensitivity to toxic dipeptide (Figure 3C), whereas strains lacking *CUP9* (*cup9Δ UBR1* and *cup9Δ ubr1Δ*) overexpressed *PTR2* (Figure 2A) and were hypersensitive to toxic dipeptide (Figure 3C).

### *Cup9p is a repressor of the PTR2 gene*

To address the mechanism of repression by Cup9p, we asked whether purified Cup9p (see Materials and methods) could selectively bind to specific regions of the *PTR2* promoter. Gel shift assays in the presence of poly-dI·dC competitor DNA were performed with labeled DNA

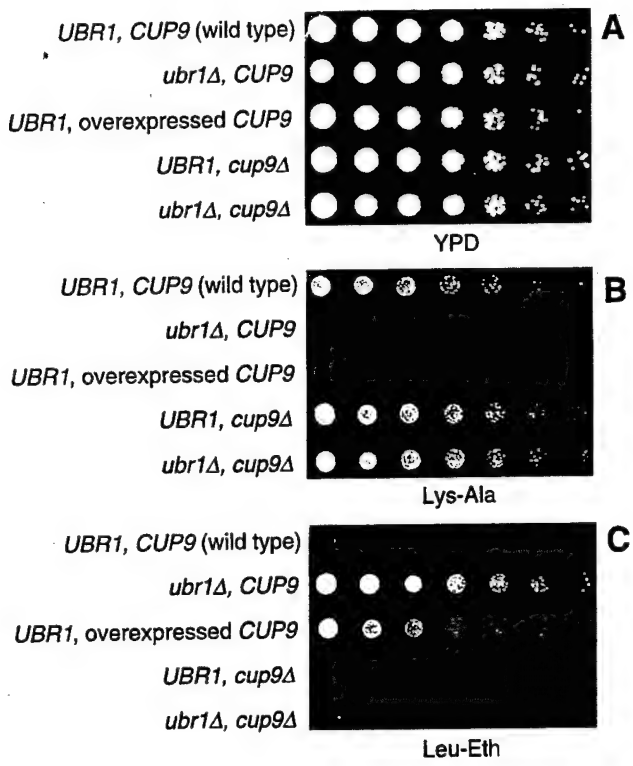



**Fig. 3.** Relative capacity for peptide import in congenic *S.cerevisiae* strains that contained, lacked or overexpressed Cup9p and/or Ubr1p. Serial dilutions of the indicated strains were deposited by a 48-pin applicator onto either rich (YPD) medium (**A**), minimal medium containing 66 μM Lys-Ala dipeptide as the sole source of lysine (**B**), or minimal medium containing both lysine (at 110 μM) and the toxic dipeptide Leu-Eth (at 55 μM) (**C**). The plates were incubated at 30°C for 1–2 days. The relevant genetic loci are shown on the left. The '*UBR1*, overexpressed *CUP9*' strain (JD52-2μ-CUP9) carried the high-copy plasmid pCB209 that expressed Cup9p from its natural promoter.

fragments of the *PTR2* promoter and $His_6$-tagged Cup9p ($Cup9p\text{-}H_6$) purified from *E.coli*. $Cup9p\text{-}H_6$ bound to a site within a distal region of the *PTR2* promoter (positions −448 to −897 relative to the inferred *CUP9* start codon), but did not bind to the proximal region of the *PTR2* promoter (positions −1 to −447) under the same conditions (Figure 2B).

The transcriptional repressor function of Cup9p was further suggested by the finding that the co-repressor complex Tup1p/Ssn6p plays a role in the control of peptide import. The Tup1p/Ssn6p complex inhibits transcription of many yeast genes through interactions with gene-specific DNA-binding repressors such as Matα2p (Chen *et al.*, 1993; Smith *et al.*, 1995), a homeodomain homolog of Cup9p. We found that most of our *sub* mutants that were not *CUP9* mutants could be complemented by low-copy plasmids bearing *SSN6* or *TUP1* (G.Turner, S.Saha and A.Varshavsky, unpublished data). In addition, deletion of *SSN6*, like deletion of *CUP9*, restored the ability of *ubr1*Δ cells to import peptides (data not shown).

### *Ubr1p-dependent degradation of Cup9p*

The fact that deletion of *CUP9* renders *PTR2* transcription independent of Ubr1p (Figure 2A), and the observation that overexpression of Ubr1p enhances the import of peptides in Cup9p-expressing strains (data not shown) suggested that the N-end rule pathway regulates peptide import by targeting Cup9p for degradation. To test this conjecture, we carried out pulse–chase experiments with a C-terminally FLAG-tagged Cup9p in *UBR1* and *ubr1*Δ cells. Cup9p-FLAG was a very short-lived protein ($t_{1/2}$ ~5 min) in *UBR1* cells (Figure 4). By contrast, Cup9p was much longer-lived ($t_{1/2}$ >30 min) in *ubr1*Δ cells (Figure 4). Degradation of Cup9p was also found to depend upon *UBC2* and *UBC4* (data not shown), in agreement with the observation that a *ubc2*Δ *ubc4*Δ double mutant failed to import dipeptides (Figure 1G). The residual instability of Cup9p in *ubr1*Δ cells (Figure 4) suggested the presence of a second, Ubr1p-independent degron; this pattern is reminiscent of another homeodomain repressor, Matα2p, which also contains at least two distinct degradation signals (Hochstrasser and Varshavsky, 1990; Chen *et al.*, 1993).

Ubr1p recognizes engineered N-end rule substrates through their destabilizing N-terminal residues (Varshavsky, 1996). To determine the N-terminal residue of Cup9p, we overexpressed and purified Cup9p-FLAG from *ubr1*Δ *S.cerevisiae* and subjected it to N-terminal sequencing. Cup9p-FLAG was found to have a blocked (presumably acetylated) N-terminus (see Materials and methods), suggesting that Ubr1p targets Cup9p through a degron distinct from the N-degron. Independent evidence for this conjecture was produced through the analysis of $GST\text{-}Cup9p\text{-}ha_2$, a fusion of glutathione transferase (GST) and C-terminally ha-tagged Cup9p. Pulse–chase analysis of $GST\text{-}Cup9p\text{-}ha_2$ revealed that the fusion protein was nearly as short-lived *in vivo* as the N-terminally unmodified Cup9p-FLAG (Figure 4, lanes b–e; compare with Figure 5, lanes b–e). If Cup9p were targeted for processive degradation through a destabilizing N-terminal residue, a preliminary proteolytic cleavage(s) of Cup9p would be required to expose such a residue (Varshavsky, 1996). In the case of $GST\text{-}Cup9p\text{-}ha_2$, this cleavage would generate a proteolytic fragment consisting of GST and an N-terminal portion of Cup9p preceding the cleavage site. Since free GST has been found to be long-lived when expressed in yeast (data not shown), accumulation of such a proteolytic fragment would be expected to accompany the degradation of $GST\text{-}Cup9p\text{-}ha_2$.

Pulse–chase analysis of $GST\text{-}Cup9p\text{-}ha_2$, using glutathione–agarose beads to isolate GST-containing proteins (see Materials and methods), indicated that the degradation of $GST\text{-}Cup9p\text{-}ha_2$ by the N-end rule pathway was not accompanied by the appearance of a fragment containing the N-terminal GST moiety (Figure 5). This finding strongly suggested that Cup9p bears an 'internal' degron recognized by Ubr1p. Additional support for this conjecture was provided by truncation analysis of a Cup9p–DHFR–myc fusion protein. These experiments indicated that the N-terminal 81 residues of the 306-residue Cup9p are dispensable for its Ubr1p-dependent degradation, strongly suggesting that the relevant degradation signal resides in the C-terminal two-thirds of Cup9p (C.Byrd, I.Davydov and A.Varshavsky, unpublished data). Although these results are fully consistent with the presence of an internal Ubr1p-dependent degron in Cup9p, there remains the less parsimonious possibility that Cup9p is degraded via *trans*-targeting (Johnson *et al.*, 1990). In this process, the two determinants of the N-degron (a destabilizing N-terminal residue and a ubiquitin-accepting internal Lys


**Fig. 4.** Degradation of Cup9p by the N-end rule pathway. *Saccharomyces cerevisiae* strains carrying pCB210, a low-copy plasmid that expressed Cup9p-FLAG from the $P_{CUP9}$ promoter, were labeled at 30°C with [$^{35}$S]methionine/cysteine for 5 min, followed by a chase for 0, 5, 15, 30 and 60 min, preparation of extracts, immunoprecipitation of Cup9p-FLAG with a monoclonal anti-FLAG antibody, SDS–PAGE and autoradiography/quantitation (see Materials and methods). Lane a, JD55 (*ubr1*Δ) cells that carried pCB200 (vector alone). Lanes b–f, JD52 (*UBR1*) (wild-type) cells that carried pCB210, expressing Cup9p-FLAG. Lanes g–k, same as lanes b–f, but with JD55 (*ubr1*Δ) cells.

residue) reside in two different polypeptides, whose interaction yields an active N-degron, leading to ubiquitinylation and degradation of the polypeptide bearing the Lys residue (Johnson *et al.*, 1990). At present, little is known about the protein ligands of Cup9p, save for the likely possibility that Cup9p interacts with the Tup1p/Ssn6p repressor complex (see above), similarly to the previously established interaction of this complex with Matα2, a homeodomain-containing homolog of Cup9p (Smith *et al.*, 1995).

## Discussion

The discovery that Ubr1p controls the import of peptides through degradation of the Cup9p repressor (Figure 6) identifies the first clear physiological function of the N-end rule pathway. The existence of mammalian, plant and bacterial homologs of the yeast Ptr2p transporter suggests that the import of peptides in these organisms may also be regulated by the N-end rule pathway.

Why was this pathway, rather than another Ub-dependent proteolytic system, recruited in the course of evolution for the regulation of peptide import? A plausible explanation is suggested by the ability of Ubr1p to bind peptides bearing destabilizing N-terminal residues (Varshavsky, 1996). Since more than half of the 20 amino acids are destabilizing in the yeast and mammalian N-end rules (Varshavsky, 1996), a significant fraction of imported peptides would be expected to compete with Cup9p for binding to Ubr1p. This competition may decrease the rate of Cup9p degradation. [Dipeptides added to a culture of *S.cerevisiae* have been shown to inhibit the N-end rule pathway (Baker and Varshavsky, 1991).] The ensuing increase in the level of Cup9p repressor would in turn decrease the production of the Ptr2p transporter, diminishing the rate of peptide import. This 'peptide-sensing' negative feedback loop would maintain the intracellular concentration of short peptides within a predetermined range—potentially a useful feature, made possible by the substrate-binding properties of Ubr1p. Experiments to verify this model are under way.

Another interesting aspect of the Ubr1p–Cup9p–Ptr2p

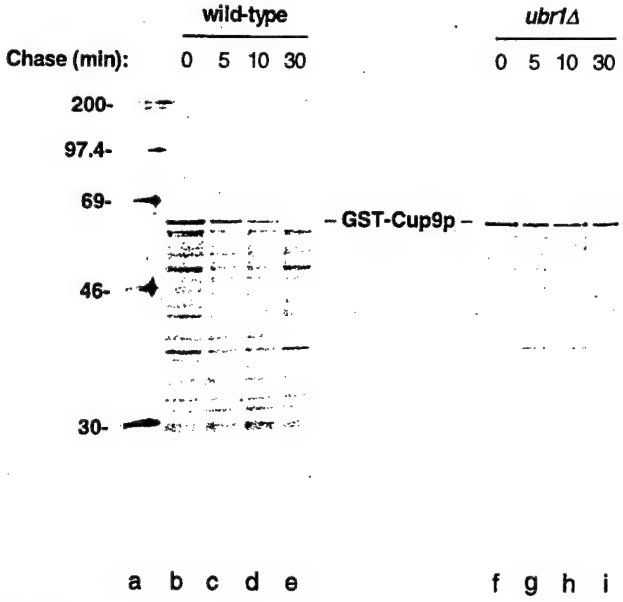



**Fig. 5.** Degradation of GST-Cup9p-ha$_2$ is not preceded by a processing that releases a GST-containing fragment. *Saccharomyces cerevisiae* strains carrying pCB120, a low-copy plasmid that expressed GST-Cup9p-ha$_2$ from the $P_{GAL1}$ promoter, were labeled at 30°C with [$^{35}$S]methionine/cysteine for 5 min, followed by a chase for 0, 5, 10 and 30 min, preparation of extracts, isolation of GST-Cup9p-ha$_2$ on glutathione–agarose beads, SDS–PAGE and autoradiography/quantitation (see Materials and methods). Lanes b–e, JD52 (*UBR1*) (wild-type) cells expressing GST-Cup9p-ha$_2$. Lanes f–i, same as lanes b–e, but with JD55 (*ubr1*Δ) cells. Lane a, molecular mass markers (in kDa).

regulatory circuit is its possible relevance to copper homeostasis. In addition to functioning as a repressor of peptide transport, Cup9p also contributes to the resistance of *S.cerevisiae* to copper toxicity during growth on lactate, a non-fermentable carbon source (Knight *et al.*, 1994). Copper homeostasis is mediated by a complex set of circuits, some of which also regulate iron metabolism (Zhou and Thiele, 1993; Ooi *et al.*, 1996). The double role of Cup9p in regulating both peptide import and copper homeostasis may signify a physiological connection between these seemingly unrelated processes.

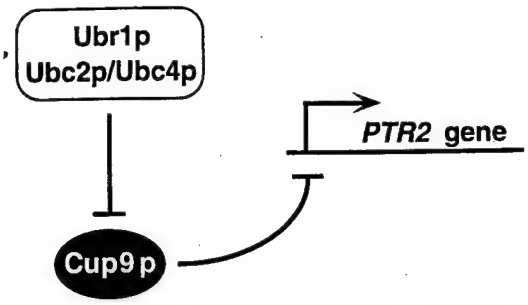



**Fig. 6.** A model for regulation of peptide import in *S.cerevisiae*. The expression of *PTR2*, which encodes a transmembrane peptide transporter, is regulated by the short-lived, homeodomain-containing transcriptional repressor Cup9p. The concentration of Cup9p is controlled in part through its degradation by the N-end rule pathway, whose targeting components include Ubr1p and either Ubc2p or Ubc4p (see the main text).

Our work adds Cup9p to the list of short-lived regulatory proteins whose degradation is mediated by the ubiquitin system. Many of these proteins are negative regulators. For example, Matα2p, a homeodomain-containing homolog of Cup9p, controls the mating type of *S.cerevisiae* (**a** or α) through repression of **a**-specific genes (Herskowitz, 1989). Matα2p appears to be constitutively short-lived in haploid cells (Hochstrasser and Varshavsky, 1990). Therefore, cessation of Matα2p synthesis upon the conversion of an α cell into an **a** cell results in rapid disappearance of Matα2p and the establishment of **a**-specific circuits (Hochstrassser, 1996). Progression of the cell cycle is also controlled by short-lived negative regulators, in particular by Sic1p, an inhibitor of CDK, the cyclin-dependent kinase (Schwob *et al.*, 1994). In this case, however, a rapid 'on-switch' is based on the phosphorylation-induced degradation of the previously stable Sic1p (King *et al.*, 1996). Whether the Cup9p repressor (Figure 6) is constitutively short-lived or whether its stability is regulated by external conditions such as, for example, different nitrogen sources, remains to be determined.

The finding that Cup9p lacks a destabilizing N-terminal residue indicates that Ubr1p, the recognition component of the N-end rule pathway, is able to target substrates bearing either internal degradation signals or N-degrons. The Gα subunit of the G protein—the other known physiological substrate of Ubr1p in *S.cerevisiae*—also lacks a destabilizing N-terminal residue (Madura and Varshavsky, 1994). Thus, a substantial fraction of naturally short-lived proteins targeted by the N-end rule pathway may bear internal degrons rather than N-degrons, a possibility that the identification of other physiological substrates of this pathway will address.

## Materials and methods

### Strains, media and genetic techniques

*Saccharomyces cerevisiae* strains were grown in rich (YPD) medium containing 2% peptone, 1% yeast extract and 2% glucose or in synthetic yeast media, containing 0.67% yeast nitrogen base without amino acids (Difco), auxotrophic nutrients at concentrations specified by Sherman *et al.* (1986) and either 2% glucose (SD medium), 2% raffinose (SM–raffinose medium) or 2% galactose (SM–galactose medium) as carbon sources. SHM–glucose medium used in halo assays was prepared according to Island *et al.* (1987) and was identical to SD medium except that it lacked methionine and contained allantoin (1 mg/ml) and yeast nitrogen base (Difco, 1.7 mg/ml) without amino acids and ammonium sulfate. *Escherichia coli* strain DH5α was used for plasmid propagation and cloning steps. For induction of the $P_{GAL1}$ promoter, cells were grown to an $A_{600}$ of 0.5–1 in SM–raffinose medium, pelleted and transferred to SM–galactose medium for 3 h.

### Halo and dilution assays

Peptide import was assayed using the halo and dilution methods. For halo assays (Island *et al.*, 1987), cells were grown to an $A_{600}$ of ~1 in SHM–glucose medium with auxotrophic supplements. Cells were pelleted by centrifugation and resuspended in water to $\sim 5\times10^7$ cells/ml. Then 0.1 ml of cell suspension was mixed with 10 ml of 0.8% noble agar (Difco) at 55°C, and spread on plates containing 20 ml of SHM medium. Sterile filter disks containing the toxic dipeptide L-leucyl-L-ethionine (Leu-Eth; 5 μmol) were placed in the middle of each plate, followed by incubation at 30°C for 1–2 days.

For dilution assays, strains were grown (under selection for plasmids) in SD medium to an $A_{600}$ of ~1. Cells from each sample ($1.5\times10^7$) were spun down and resuspended in 1 ml of water. The samples were serially diluted by 4-fold in a microtiter plate (150 μl/well) to generate eight different initial concentrations of cells that ranged from $1.5\times10^7$ to 915 cells/ml. The suspensions were spotted to various media, using a 48-pin applicator. The plates were incubated at 30°C for 1–2 days.

Leu-Eth was synthesized using standard methods of organic chemistry. Briefly, L-ethionine methyl ester (Eth-OMe) was produced from L-ethionine and methanol. Eth-OMe was then coupled with *N-tert*-butoxy-carbonyl-L-leucine-*p*-nitrophenyl ester (N-*t*-BOC-L-leucine-PNP), yielding N-*t*-BOC-L-leucyl-L-ethionine methyl ester, which was purified by flash chromatography on a silica column. N-*t*-BOC-L-leucyl-L-ethionine methyl ester was then converted into N-*t*-BOC-L-leucyl-L-ethionine by treatment with KOH. N-*t*-BOC-L-leucyl-L-ethionine was deprotected with trifluoroacetic acid, yielding Leu-Eth.

### A screen for import-competent mutants in the ubr1Δ background

Thirty 5-ml cultures of JD83-1A (*leu2-3 ubr1Δ*), auxotrophic for leucine, were grown to an $A_{600}$ of ~1, and $\sim1.5\times10^7$ cells from each culture were plated onto SD medium lacking leucine and histidine but containing the leucyl-histidine dipeptide at 0.23 μM. Plates were incubated at 30°C for 2 days, selecting for mutants able to grow on this medium. Of the 320 sub mutants ('suppressors of a block to peptide import in *ubr1Δ*') thus obtained, 199 were found by complementation tests to be clearly recessive. Among these, 101 mutants belonged to one complementation group, termed *sub1*.

### Isolation of the CUP9 (SUB1) gene

A 50 ml culture of *S.cerevisiae* CBY15 (*ubr1Δ sub1-1*) was grown to an $A_{600}$ of ~1 in SHM–glucose. Cells were made competent with lithium acetate (Ausubel *et al.*, 1992), and 25 0.1-ml samples ($\sim7\times10^7$ cells/ml) were transformed with a yeast genomic DNA library (American Type Culture Collection; #77164) in the *TRP1*, *CEN6*-based vector pRS200 (Sikorski and Hieter, 1989). The cells were pelleted, and each sample was resuspended in 1.5 ml of SHM–glucose and incubated at 30°C for 3 h. The cells were pelleted again, each sample was resuspended in 0.1 ml of water, added to 10 ml of 0.8% Noble agar at 55°C, and spread onto SHM–glucose plates lacking Trp and containing Leu-Eth at 37 μM. Of the $\sim2.5\times10^4$ $Trp^+$ transformants plated, 14 could grow in the presence of Leu-Eth. Of these, two yielded the library-derived plasmids, pSUB1-1 and pSUB1-6, that complemented the *sub1* mutation of CBY15. The insert of pSUB1-6 contained the insert of pSUB1-1. Partial sequencing of the ~8.2 kb insert of pSUB1-1 identified it as a region of the *S.cerevisiae* chromosome XVI. Deletion analysis (not shown) localized the complementing activity to an ~2.8 kb *Hin*dIII–*Kpn*I fragment, whose sequence revealed the presence of two ORFs (*CUP9* and SCYPL178W). Further deletion analysis (not shown) localized the complementing activity to the previously isolated (Knight *et al.*, 1994) *CUP9* gene. Verification that *CUP9* and *SUB1* were one and there same gene was carried out as described in Results.

### Construction of null cup9 mutants

The ~2.8 kb *Hin*dIII–*Kpn*I genomic DNA fragment containing *CUP9* was ligated to *Hin*dIII–*Kpn*I-cut pRS306ΔSpeI, yielding pCB117. pRS306ΔSpeI was derived from the *URA3*-bearing pRS306 (Sikorski and Hieter, 1989) through elimination of its *Spe*I site. The *CUP9* ORF was disrupted by inserting an ~2 kb, blunt-ended, *LEU2*-containing *Sal*I fragment from pJJ283 into the *Spe*I site of *CUP9*, yielding pCB119, in which the *LEU2* and *CUP9* ORFs were oriented in opposite directions. An ~5 kb *Hin*dIII–*Pvu*II fragment of pCB119 containing the *cup9::LEU2*

**Table I.** *Saccharomyces cerevisiae* strains used in this study

| Strain | Genotype | References |
|---|---|---|
| DF5 | *MATa/MATα trp1-1/trp1-1 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801 gal/gal* | Finley *et al.* (1987) |
| EJ1 | *MATα ubc5Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal* | Derivative of DF5[a] |
| EJ2 | *MATα ubc1Δ::URA3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal* | Derivative of DF5[a] |
| EJ3 | *MATα ubc4Δ::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal* | Derivative of DF5[a] |
| EJ10 | *MATα ubc4Δ::HIS3 ubc5Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal* | Derivative of DF5[a] |
| EJY102 | *MATα ubc2Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal* | Derivative of DF5[a] |
| EJY105 | *MATa ubc2Δ::LEU2 ubc4Δ::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal* | Derivative of DF5[a] |
| EJY106 | *MATα ubc2Δ::LEU2 ubc4Δ::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal* | Derivative of DF5[a] |
| MHY495 | *MATα ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of DF5[b] |
| MHY503 | *MATα ubc4-Δ1::HIS3 ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of DF5[b] |
| MHY507 | *MATα ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of DF5[b] |
| MHY550 | *MATα ubc4-Δ2::TRP1 ubc5-Δ1::LEU2 ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of DF5[b] |
| MHY552 | *MATα ubc6-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of DF5[b] |
| MHY554 | *MATα ubc4-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of DF5[b] |
| MHY557 | MATα ubc4-Δ1::HIS3 ubc6-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 | Derivative of DF5[b] |
| MHY570 | *MATα ubc4-Δ2::TRP1 ubc5-Δ1::LEU2 ubc6-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of DF5[b] |
| MHY599 | *MATα pas2 trp ura3-52 ade1 leu2-3* | Chen *et al.* (1993) |
| MHY601 | *MATα ubc8::URA3 trp1-1 ura3-1 ade2-1 his3-11 leu2-3,112 can1-100* | Chen *et al.* (1993) |
| YWO55 | *MATα ubc9-1 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of DF5[c] |
| YPH500 | *MATα trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801* | Sikorski and Hieter (1989) |
| BBY67 | *MATα ubc2Δ::LEU2 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801* | Derivative of YPH500[d] |
| KM207-1 | *MATα ate1Δ::TRP1 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801* | Derivative of YPH500[e] |
| JD34 | *MATa ubc2Δ::URA3 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801* | Derivative of YPH500[f] |
| JD51 | *MATa/MATα trp1-Δ63/trp1-Δ63 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801* | Dohmen *et al.* (1995) |
| JD52 | *MATa trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Johnson *et al.* (1995) |
| JD53 | *MATα trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Dohmen *et al.* (1995) |
| JD55 | *MATa ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Madura and Varshavsky (1994) |
| JD83-1A | *MATα ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of JD51[f] |
| RJD549 | *MATa cdc34-1 trp1 ura3-52 leu2-3* | R.Deshaies[g] |
| RJD795 | *MATa cdc34-2 trp1 ura3-52 leu2-3* | R.Deshaies[g] |
| SGY6 | *MATα nta1Δ::TRP1 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-3,112 lys2-801* | S.Grigoryev[h] |
| CBY4 | *MATα ubc2Δ::URA3 ubc5Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal* | Derivative of EJ1 |
| CBY5 | *MATα ubc2Δ::LEU2 ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of MHY495 |
| CBY6 | *MATα ubc2Δ::URA3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of MHY507 |
| CBY7 | *MATα ubc2Δ::LEU2 ubc8::URA3 trp1-1 ura3-1 ade2-1 his3-11 leu2-3,112 can1-100* | Derivative of MHY601 |
| CBY8 | *MATα ubc2Δ::LEU2 ubc1Δ::URA3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal* | Derivative of EJ2 |
| CBY9 | *MATα ubc2Δ::LEU2 ubc9-1 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of YWO55 |
| CBY10 | *MATα ubc2Δ::LEU2 pas2 trp ura3-52 ade1 leu2-3* | Derivative of MHY599 |
| CBY11 | *MATa ubc2Δ::LEU2 cdc34-1 trp1 ura3-52 leu2-3* | Derivative of RJD549 |
| CBY15 | *MATα sub1-1 ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of JD83-1A |
| CBY16 | *MATa cup9::LEU2 ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of JD55 |
| CBY17 | *MATa cup9::LEU2 ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of JD55 |
| CBY19 | *MATa cup9::LEU2 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801* | Derivative of JD52 |
| CBY23 | *MATa/MATα sub1-1/cup9Δ::LEU2 ubr1Δ::HIS3/ubr1Δ::HIS3 trp1-Δ63/trp1-Δ63 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801* | Produced by mating CBY15 and CBY16 |
| CBY24 | *MATa/MATα sub1-1/cup9Δ::LEU2 ubr1Δ::HIS3/ubr1Δ::HIS3 trp1-Δ63/trp1-Δ63 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801* | Produced by mating CBY15 and CBY17 |

[a]Johnson *et al.* (1992, 1995). A gift from E.Johnson, the Rockefeller University, New York, NY 10021-6399, USA.
[b]Chen *et al.* (1993). A gift from M.Hochstrasser, Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA.
[c]A gift from S.Jentsch, ZMBH, Universität Heidelberg, 69120 Heidelberg, Germany.
[d]Dohmen *et al.* (1990).
[e]A gift from K.Madura, Department of Biochemistry, UMDNJ-Johnson Medical School, Piscataway, NJ 08854, USA.
[f]A gift from J.Dohmen, Heinrich-Heine-Universität, Institut für Mikrobiologie, 40225 Düsseldorf, Germany.
[g]A gift from R.Deshaies, Division of Biology, Caltech, Pasadena, CA 91125, USA.
[h]A gift from S.Grigoryev, Department of Biology, University of Massachusetts, Amherst, MA 01003, USA.

disruption allele was used to replace the wild-type *CUP9* alleles of JD52 (wild-type) and JD55 (*ubr1Δ*) by homologous recombination (Rothstein, 1991), generating strains CBY19 and CBY17, respectively.

### *CUP9-expressing plasmids*

The plasmid pCB116, which expressed Cup9p-FLAG from the $P_{GAL1}$ promoter, was constructed by subcloning an ~1 kb *Bam*HI–*Eco*RI fragment containing the *CUP9-FLAG* ORF into the *Bam*HI–*Eco*RI site(s) of p416GAL1 (Mumberg *et al.*, 1994). The *CUP9-FLAG*-containing fragment was constructed by PCR amplification of the *CUP9* ORF of plasmid pCB111 using primers PCB1 (5′-CGCGGATCCGAATAGTTACATTCGAAGATG-3′) and PCB6 (5′-CCGGAATTCTCATTTATCATCATCGTCTTTGTAATCATTCATATCAGGGTTGGATAG-3′), resulting in the addition of the 8-residue FLAG epitope, DYKDDDDK, to the C-terminus of Cup9p. pCB111 was constructed by subcloning the ~2.8 kb *Hin*dIII–*Kpn*I fragment of the *CUP9*-containing pSUB1-1 (see above) into *Hin*dIII–*Kpn*I-cut pRS316 (Sikorski and Hieter, 1989). pCB202 was constructed by subcloning an ~1 kb fragment containing the $P_{CUP9}$ promoter into *Hin*dIII–*Bam*HI-cut pCB201. [The ~1 kb fragment was produced by PCR from pCB111 using primers PCB8

(5′-GTGTTAGTAAGCTTGTAAAGGAATGCACGTATT-3′) and PCB9 (5′-CCCGCGGATCCGCATGCAACTATTCTCGAAGGTTGT-3′).] pCB200 (*ARS-CEN, LEU2*) and pCB201 (2μ, *LEU2*) were constructed by replacing the 517 bp *Sca*I–*Eco*RI fragment of pBR322 (Ausubel *et al.*, 1992) with, respectively, the 3822 bp *Sca*I–*Nae*I fragment of pRS415 (Sikorski and Hieter, 1989) and the 4650 bp *Sca*I–*Nae*I fragment of pRS425 (Christianson *et al.*, 1992).

The plasmid pCB209 (2μ, *LEU2*), which expressed *CUP9* from the $P_{CUP9}$ promoter, was constructed by replacing the *Sph*I–*Sal*I fragment of pCB202 with an ~1 kb fragment containing the *CUP9* ORF that was produced by PCR from pCB111, using primers PCB10 (5′-CCCGCGGATCCGCATGCGAAGATGAATTATAACTGC-3′) and PCB12 (5′-CCCGCGCGGTCGACCTCAATTCATATCAGGGTTGGATAG-3′). pCB210 (*ARS-CEN, LEU2*) that expressed Cup9p-FLAG from the $P_{CUP9}$ promoter was constructed by replacing the *Sph*I–*Sal*I fragment of pCB202 with an ~1 kb fragment containing the *CUP9-FLAG* ORF, which was produced from pCB111 using primers PCB10 (see above) and PCB13 (5′-CCCGCGCGGTCGACCTCATTTATCATCATCGTCTTTGTA-ATCATTCATATCAGGGTTGGATAG-3′), yielding pCB211. The ~2 kb *Hin*dIII–*Sal*I fragment of pCB211 containing $P_{CUP9}$ and the *CUP9-FLAG* ORF was subcloned into pCB200, yielding pCB210.

Plasmid pCB120, expressing GST-Cup9p-$ha_2$ from the $P_{GAL1}$ promoter, was constructed by subcloning the ~1.6 kb *Xba*I–*Eco*RI fragment, containing the *GST-CUP9-*$ha_2$ ORF, into the *Xba*I–*Eco*RI site(s) of p416GAL1. The *Xba*I–*Eco*RI fragment was produced by PCR amplification of the *GST-CUP9-*$ha_2$ ORF of pGEX-2T–CUP9-$ha_2$, using the primers PCB3 (5′-CCGGAATTCTCAAGCGTAATCTGGAACATC-GTATGGGTAAGCGTAATCTGGAACATCGTATGGGTAATTCATA-TCAGGGTTGGATAG-3′) and PCB5 (5′-TGCTCTAGAACAGT-ATTCATGTCCCCTATA-3′). pGEX-2T–CUP9-$ha_2$ was constructed by subcloning an ~1 kb fragment containing the *CUP9-*$ha_2$ ORF into the *Bam*HI–*Eco*RI site(s) of pGEX-2T (Pharmacia), resulting in an in-frame fusion of the sequence encoding 26 kDa glutathione *S*-transferase (GST) domain of *Saccharomyces japonicum* (Smith and Johnson, 1988) to the second codon of *CUP9*. The *CUP9-*$ha_2$-containing fragment was produced by PCR amplification of the *CUP9* ORF of pCB111 using the primers PCB3 (see above) and PCB4 (5′-CGCGGATCCAATTAT-AACTGCGAAATACAAAAC-3′). This step added to the C-terminus of Cup9p a sequence encoding a tandem repeat of the 9-residue sequence YPYDVPDYA, derived from hemagglutinin (ha) of influenza virus.

### Northern hybridization

RNA was isolated from *S.cerevisiae* as described (Schmitt *et al.*, 1990). Electrophoresis of the RNA samples was carried out on a formaldehyde RNA gel (Ausubel *et al.*, 1992). An ~50-μg RNA sample was loaded on a 1% agarose gel containing 1× MOPS buffer, 0.74% (v/v) formaldehyde, 1.9 mg/ml iodoacetamide and 0.5 μg/ml ethidium bromide. Electrophoresis was carried out in 1× MOPS buffer at 5 V/cm. RNA was transferred to BrightStar-Plus membrane (Ambion) using TurboBlotter (Schleicher & Schuell) and Ambion RNA transfer buffer. RNA was crosslinked to the air-dried membranes using 254 nm light (Ausubel *et al.*, 1992).

DNA probes were prepared by the random priming method (Ausubel *et al.*, 1992) using [$^{32}$P]dCTP and a DNA labeling kit (Pharmacia). Hybridization was carried out for 8–16 h at 42°C in Prehybridization/Hybridization Solution (Ambion). Filters were washed according to the manufacturer's protocol and subjected to autoradiography.

### Gel shift assay

PCR was used to extend the Cup9p ORF with a sequence encoding Ser-Gly-Gly-Thr-$His_6$, yielding Cup9p-$H_6$, and to engineer flanking restriction sites (*Nde*I and *Bam*HI) for insertion into pET-11c (Novagen). Cup9p-$H_6$ was overexpressed in *E.coli* BL21 (DE3) (Novagen) (Ausubel *et al.*, 1992) and purified on a 3 ml Ni-NTA column (Qiagen), using a linear gradient of imidazole. Cup9p-$H_6$ eluted at ~0.25 M imidazole (~90% pure at this step); it was dialyzed at 4°C against 10% glycerol, 0.1 M KCl, 1 mM EDTA, 0.5 mM dithiothreitol, 20 mM HEPES, pH 7.9, and then snap-frozen in multiple samples in liquid $N_2$, and stored at –80°C. The proximal (–1 to –447) and distal (–448 to –897) *PTR2* promoter probes for the gel shift assay were constructed by PCR amplification in the presence of [α-$^{32}$P]dCTP, and were purified using spin columns (Qiagen). The gel shift reactions (20 μl) contained 50 μg/ml poly-dI·dC (Pharmacia); ~1.5 μg/ml (500 c.p.m.) DNA probe; 1 mg/ml acetylated serum albumin (New England Biolabs) and either 0.1, 0.5 or 1 μg/ml of Cup9p-$H_6$ in 10% glycerol, 0.1 M KCl, 2.5 mM $MgCl_2$, 1 mM EDTA, 20 mM HEPES, pH 7.9. The samples were incubated for 30 min at room temperature, then loaded onto a 4% polyacrylamide gel (40:1, acrylamide:*bis*-acrylamide) in 0.5× TBE (Ausubel *et al.*, 1992), and electrophoresed at 10 V/cm for 3 h at 4°C, followed by autoradiography.

### Pulse–chase analysis of Cup9p

One hundred ml cultures of *S.cerevisiae* JD52 (*UBR1*) and JD55 (*ubr1*Δ) carrying either pCB210 (expressing Cup9p-FLAG from the *CUP9* promoter) or pCB200 (vector alone) were grown to an $A_{600}$ of ~1 in SD(–Leu) medium. Cells (50 $A_{600}$ units total) from each of the four cultures were gently pelleted by centrifugation, washed with 5 ml of SD(–Leu), pelleted again, resuspended in 2 ml of SD(–Leu), and incubated at 30°C for 10 min. Each sample was labeled for 5 min with 1.4 mCi of [$^{35}$S]methionine/cysteine (EXPRESS, New England Nuclear) at 30°C, followed by pelleting in a microfuge for ~15 s. The cells were resuspended in 2.6 ml of SD(–Leu), 5 mM L-methionine, 5 mM L-cysteine, and incubated at 30°C. Samples of 0.5 ml were withdrawn during the incubation, pelleted and resuspended in 0.15 ml of 0.5 M NaCl-Lysis Buffer (1% Triton X-100, 0.5 M NaCl, 5 mM EDTA, 50 mM Na-HEPES, pH 7.5) containing a mixture of protease inhibitors (Ghislain *et al.*, 1996). Glass beads (0.5 mm) were added, and cells were disrupted by vortexing (six times, for 30 s each, with 1 min incubations on ice in between), followed by the adjustment of NaCl concentration to 0.15 M through the addition of 75 mM NaCl-Lysis Buffer, further vortexing for 30 s, and centrifugation at 12 000 *g* for 10 min. The volumes of supernatants were adjusted to equalize the amounts of 10% trichloroacetic acid-insoluble $^{35}$S. Cup9p-FLAG was immunoprecipitated by the addition of 20 μl of the monoclonal anti-FLAG M2 antibody conjugated to agarose beads (Kodak). Suspensions were incubated at 0°C for 1 h, with rotation, then centrifuged at 12 000 *g* for 30 s, and washed four times with 0.8 ml of 0.15 M NaCl-Lysis Buffer. The pellets were resuspended in SDS-sample buffer, heated at 100°C for 3 min, and subjected to SDS–12% PAGE, followed by autoradiography and quantitation using a PhosphorImager (Molecular Dynamics).

Pulse–chase analysis of GST-Cup9p-$ha_2$ was carried out as described by Bartel *et al.* (1990). Approximately 10 $A_{600}$ units of galactose-induced cells were labeled for 5 min with 0.3 mCi of [$^{35}$S]EXPRESS in 400 μl SM–galactose (–Ura) at 30°C. The cells were then transferred to microfuge tubes, pelleted and resuspended in 500 μl of SD (–Ura), 5 mM L-methionine, 5 mM L-cysteine. Samples of 0.1 ml were withdrawn during the incubation, pelleted and lysed as above. The $^{35}$S-labeled GST-Cup9p-$ha_2$ was purified using glutathione–agarose beads (Sigma) which had been blocked with bovine serum albumin (BSA; 10 mg/ml). Twenty μl of glutathione–agarose beads were added to each sample and the suspensions were incubated at 0°C for 60 min, with rotation, followed by washes and electrophoretic analyses as described for Cup9p-FLAG.

### Purification and N-terminal sequencing of Cup9p-FLAG

Four 2-l cultures of JD55 (*ubr1*Δ) carrying pCB116 that expressed Cup9p-FLAG from the *GAL1* promoter were grown under selection in SM–raffinose to an $A_{600}$ of ~0.8, followed by transfer to SM–galactose and incubation at 30°C for 3 h. Longer induction times resulted in a Cup9p-mediated cytotoxicity and lower yields of Cup9p-FLAG. The cells (~$1\times10^{11}$) were harvested and lysed at 4°C as described by Burgers (1995). The extract was fractionated by precipitation with 0.4% Polymin P (Sigma) and then further by precipitation with 48% saturated ammonium sulfate. The pellet was dissolved in 3 ml of TBS buffer (0.15 M NaCl, 50 mM Tris–HCl, pH 7.5), and passed through Sephadex G-25 in TBS. The resulting sample (8.3 ml, ~70 mg/ml of protein) was applied to a column (1 ml) of the monoclonal anti-FLAG M2 antibody (Kodak). The column was washed three times with 10 ml of TBS, and Cup9p-FLAG was eluted by the addition of five 1-ml samples of TBS containing, respectively, 50, 100, 100, 200 and 200 μg/ml of the FLAG peptide (Kodak). Peak Cup9p-FLAG fractions (detected by immunoblotting) were concentrated by partial lyophilization, followed by precipitation with methanol (Wessel and Flügge, 1984) in the presence of human insulin (Sigma, 0.3 mg/ml) as a carrier. The resulting sample was fractionated by SDS–12% PAGE and electroblotted onto Pro-Blot membrane (Perkin-Elmer). After a brief staining with Coomassie, the band of the 37 kDa Cup9p-FLAG (~15 pmol) was excised from the membrane. Half of the sample was used to determine the amino acid composition; the other half was subjected to N-terminal sequencing for seven cycles, using the Applied Biosystems 476A protein sequencer at the Caltech Microchemistry Facility.

## Acknowledgements

We thank R.Deshaies, R.J.Dohmen, S.Grigoryev, M.Hochstrasser, S.Jentsch, E.S.Johnson, K.Madura, I.Ota and C.Trotta for the gifts of

strains and plasmids: S.Carter for his guidance in the synthesis of Leu-Eth; N.Johnsson, R.Deshaies, R.J.Dohmen and A.Webster for helpful discussions; L.Peck, Y.T.Kwon, A.Webster and F.Du for comments on the manuscript; S.Saha for help in overexpressing Cup9p; and N.Riley for technical assistance. This work was supported by grants to A.V. from the National Institutes of Health (DK39520 and GM31530).

## References

Alagramam,K., Naider,F. and Becker,J.M. (1995) A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. *Mol. Microbiol.*, **15**, 225–234.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Smith,J.A., Seidman,J.G. and Struhl,K. (1992) *Current Protocols in Molecular Biology*. Wiley-Interscience, New York.

Bachmair,A. and Varshavsky,A. (1989) The degradation signal in a short-lived protein. *Cell*, **56**, 1019–1032.

Bachmair,A., Finley,D. and Varshavsky,A. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science*, **234**, 179–186.

Baker,R.T. and Varshavsky,A. (1991) Inhibition of the N-end rule pathway in living cells. *Proc. Natl Acad. Sci. USA*, **87**, 2374–2378.

Baker,R.T. and Varshavsky,A. (1995) Yeast N-terminal amidase: a new enzyme and component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 12065–12074.

Balzi,E., Choder,M., Chen,W.A., Varshavsky,A. and Goffeau,A. (1990) Cloning and functional analysis of the arginyl-tRNA–protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **265**, 7464–7471.

Bartel,B. (1990) *Molecular Genetics of the Ubiquitin System: the Ubiquitin Fusion Proteins and Proteolytic Targeting Mechanisms*. PhD Thesis, M.I.T., Cambridge, MA, USA.

Bartel,B., Wünning,I. and Varshavsky,A. (1990) The recognition component of the N-end rule pathway. *EMBO J.*, **9**, 3179–3189.

Burgers,P.M.J. (1995) Preparation of extracts from yeast and avoidance of proteolysis. *Methods Mol. Cell. Biol.*, **5**, 330–335.

Chau,V., Tobias,J.W., Bachmair,A., Marriott,D., Ecker,D.J., Gonda,D.K. and Varshavsky,A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, **243**, 1576–1583.

Chen,P., Johnson,P., Sommer,T., Jentsch,S. and Hochstrasser,M. (1993) Multiple ubiquitin-conjugating enzymes participate in the *in vivo* degradation of the yeast MATα2 repressor. *Cell*, **74**, 357–369.

Christianson,T.W., Sikorski,R.S., Dante,M. and Hieter,P. (1992) Multifunctional yeast high-copy-number shuttle vectors. *Gene*, **110**, 119–122.

deGroot,R.J., Rümenapf,T., Kuhn,R.J. and Strauss,J.H. (1991) Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc. Natl Acad. Sci. USA*, **88**, 8967–8971.

Dohmen,R.J., Stappen,R., McGrath,J.P., Forrová,H., Kolarov,J., Goffeau,A. and Varshavsky,A. (1995) An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. *J. Biol. Chem.*, **270**, 18099–18109.

Finley,D., Özkaynak,E. and Varshavsky,A. (1987) The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation and other stresses. *Cell*, **48**, 1035–1046.

Ghislain,M., Dohmen,R.J., Lévy,F. and Varshavsky,A. (1996) Cdc48p interacts with Ufd3p, a WD-repeat protein required for ubiquitin-dependent proteolysis in *Saccharomyces cerevisiae*. *EMBO J.*, **15**, 4884–4899.

Herskowitz,I. (1989) A regulatory hierarchy for cell specialization in yeast. *Nature*, **342**, 749–757.

Hilt,W. and Wolf,D.H. (1996) Proteasomes: destruction as a programme. *Trends Biochem. Sci.*, **21**, 96–102.

Hochstrasser,M. (1996) Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.*, **30**, 405–439.

Hochstrasser,M. and Varshavsky,A. (1990) *In vivo* degradation of a transcriptional regulator: the yeast α2 repressor. *Cell*, **61**, 697–708.

Island,M.D., Naider,F. and Becker,J.M. (1987) Regulation of dipeptide transport in *S.cerevisiae* by micromolar amino acid concentrations. *J. Bacteriol.*, **169**, 2132–2136.

Jentsch,S. (1992) The ubiquitin-conjugating system. *Annu. Rev. Genet.*, **26**, 179–207.

Jentsch,S. and Schlenker,S. (1995) Selective protein degradation: a journey's end within the proteasome. *Cell*, **82**, 881–884.

Johnson,E.S., Gonda,D.K. and Varshavsky,A. (1990) *Cis–trans* recognition and subunit-specific degradation of short-lived proteins. *Nature*, **346**, 287–291.

Johnson,E.S., Bartel,B., Seufert,W. and Varshavsky,A. (1992) Ubiquitin as a degradation signal. *EMBO J.*, **11**, 497–505.

Johnson,E.S., Ma,P.C.M., Ota,I.M. and Varshavsky,A. (1995) A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.*, **270**, 17442–17456.

King,R.W., Deshaies,R.J., Peters,J.M. and Kirschner,M.W. (1996) How proteolysis drives the cell cycle. *Science*, **274**, 1652–1659.

Knight,S.A.B., Tamai,K.T., Kosman,D.J. and Thiele,D.J. (1994) Identification and analysis of a *Saccharomyces cerevisiae* copper homeostasis gene encoding a homeodomain protein. *Mol. Cell. Biol.*, **14**, 7792–7804.

Madura,K. and Varshavsky,A. (1994) Degradation of Gα by the N-end rule pathway. *Science*, **265**, 1454–1458.

Madura,K., Dohmen,R.J. and Varshavsky,A. (1993) N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.*, **268**, 12046–12054.

Mumberg,G., Müller,R. and Funk,M. (1994) Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.*, **22**, 5767–5768.

Ooi,C.E., Rabinovich,E., Dancis,A., Bonifacino,J.S. and Klausner,R.D. (1996) Copper-dependent degradation of the *Saccharomyces cerevisiae* plasma membrane copper transporter Ctr1p in the apparent absence of endocytosis. *EMBO J.*, **15**, 3515–3523.

Rothstein,R. (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol.*, **194**, 281–301.

Rubin,D.M., van Nocker,S., Glickman,M., Coux,O., Wefes,I., Sadis,S., Fu,H., Goldberg,A., Vierstra,R. and Finley,D. (1997) ATPase and ubiquitin-binding proteins of the yeast proteasome. *Mol. Biol. Rep.*, **24**, 17–26.

Schmitt,M.E., Brown,T.A. and Trumpower,B.L. (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **18**, 3091–3092.

Schwob,E., Bohm,T., Mendenhall,M.D. and Nasmyth,K. (1994) The B-type cyclin kinase inhibitor p40 (Sic1) controls the G1 to S transition in *Saccharomyces cerevisiae*. *Cell*, **79**, 233–244.

Sherman,F., Fink,G.R. and Hicks,J.B. (1986) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sikorski,R.S. and Hieter,P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *S. cerevisiae*. *Genetics*, **122**, 19–27.

Smith,D.B. and Johnson,K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*, **67**, 31–38.

Smith,R.L., Redd,M.J. and Johnson,A.D. (1995) The tetratricopeptide repeats of Ssn6 interact with the homeodomain of α2. *Genes Dev.*, **9**, 2903–2910.

Varshavsky,A. (1996) The N-end rule: functions, mysteries, uses. *Proc. Natl Acad. Sci. USA*, **93**, 12142–12149.

Wessel,D. and Flügge,U.I. (1984) A method for quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.*, **138**, 141–143.

Wolberger,C. (1996) Homeodomain interactions. *Curr. Opin. Struct. Biol.*, **6**, 62–68.

Zhou,P. and Thiele,D.J. (1993) Copper and gene regulation in yeast. *Biofactors*, **4**, 105–115.

*Received August 13, 1997; revised October 13, 1997; accepted October 14, 1997*

REVIEW

# The N-end rule pathway of protein degradation

**Alexander Varshavsky***

*Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA*

**The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct versions of the N-end rule operate in all organisms examined, from mammals to fungi and bacteria. In eukaryotes, the N-end rule pathway is a part of the ubiquitin system. Ubiquitin is a 76-residue protein whose covalent conjugation to other proteins plays a role in many biological processes, including cell growth and differentiation. I discuss the current understanding of the N-end rule pathway.**

## Introduction

Many intracellular proteins are metabolically unstable, or can become unstable during their lifetime in a cell. The functions of intracellular proteolysis include the elimination of abnormal proteins, the maintenance of amino acid pools in cells affected by stresses such as starvation, and the generation of protein fragments that act as hormones, antigens or other effectors. Yet another function of proteolytic pathways is the selective destruction of proteins whose concentrations must vary with time, and alterations in the state of a cell. Metabolic instability is a property of many regulatory proteins. A short *in vivo* half-life[a] of a regulator provides a way to generate its spatial gradients and allows for rapid adjustments of its concentration (or subunit composition) through changes in the rate of its synthesis. A protein can also be conditionally unstable, i.e., long-lived or short-lived depending on the state of a cell. Conditionally short-lived regulators are often deployed as components of control circuits. One example is cyclins—a family of related proteins whose destruction at specific stages of the cell cycle regulates cell division and growth (Murray & Hunt 1993). In addition, many proteins are long-lived as components of larger complexes such as ribosomes and oligomeric proteins, but are metabolically unstable as free subunits. The short *in vivo* lifetimes of free subunits allow for a less stringent control over the relative rates of their synthesis, since a subunit produced in excess would not accumulate to a significant level.

Features of proteins that confer metabolic instability are called degradation signals, or degrons (Varshavsky 1991). The essential component of one degradation signal—the first to be discovered—is a destabilizing N-terminal residue of a protein (Bachmair *et al.* 1986; Varshavsky 1992, 1996a). This signal is called the N-degron. A set of N-degrons containing different destabilizing residues yields a rule, termed the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue (Table 1 and Fig. 1). The N-end rule pathway is present in all organisms examined, including the bacterium *Escherichia coli* (Tobias *et al.* 1991; Shrader *et al.* 1993), the yeast (fungus) *Saccharomyces cerevisiae* (Bachmair & Varshavsky 1989), and mammalian cells (Gonda *et al.* 1989; Lévy *et al.* 1996) (Fig. 1).

The N-end rule was encountered in experiments that explored the metabolic fate of a fusion between Ub and a reporter protein such as *E. coli* β-galactosidase (βgal) in *S. cerevisiae* (Bachmair *et al.* 1986). In yeast and other eukaryotes, Ub-X-βgal is cleaved, cotranslationally or nearly so, by Ub-specific processing proteases at the Ub-βgal junction. This cleavage takes place regardless

* *Correspondence*: E-mail: varshavskya@starbase1.caltech.edu

[a] The *in vivo* degradation of many short-lived proteins, including the engineered N-end rule substrates, deviates from first-order kinetics (Baker & Varshavsky 1991). Therefore the term 'half-life', if applied to an entire decay curve, is a useful but often crude approximation. A more rigorous terminology for describing nonexponential decay was proposed by Lévy *et al.* (1996).

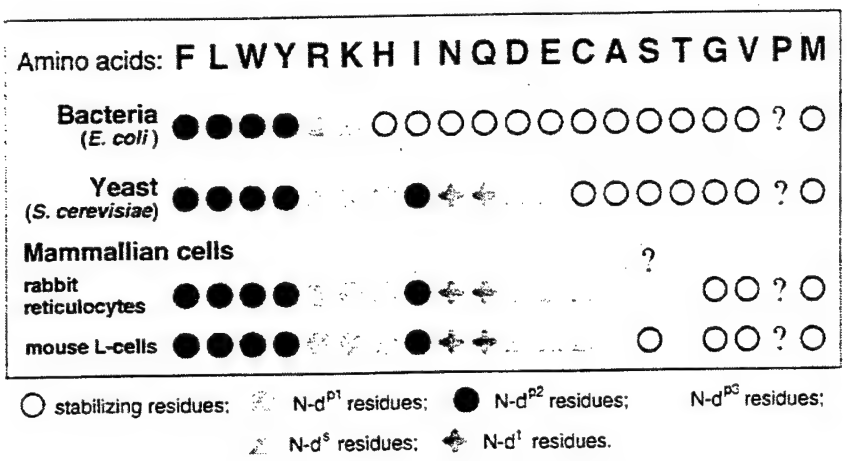



**Figure 1** Comparison of eukaryotic and bacterial N-end rules. Open circles denote stabilizing residues. Purple and red circles denote, respectively, type 1 and type 2 primary destabilizing residues. Blue triangles denote secondary destabilizing residues. Green crosses denote tertiary destabilizing residues (Varshavsky 1996a). A question mark at Pro indicates its uncertain status (see the main text). A question mark above Ser indicates its uncertain status in the reticulocyte N-end rule (Lévy *et al.* 1996). Single-letter abbreviations for amino acids: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

of the identity of the residue X at the C-terminal side of the cleavage site, proline being the single exception. By allowing a bypass of the normal N-terminal processing of a newly formed protein, this finding (Fig. 2A) yielded an *in vivo* method for generating different residues at the N-termini of otherwise identical proteins—a technical advance that led to the N-end rule (Varshavsky 1992, 1996a).

In eukaryotes, the N-degron comprises at least two determinants: a destabilizing N-terminal residue and an internal lysine (or lysines) of a substrate (Fig. 2B) (Bachmair & Varshavsky 1989; Johnson *et al.* 1990; Dohmen *et al.* 1994). The Lys residue is the site of formation of a multiubiquitin chain (Chau *et al.* 1989). Ubiquitin (Ub) is a 76-residue protein whose covalent conjugation to other proteins is involved in a multitude of processes—cell growth and differentiation, signal transduction, DNA repair, transmembrane traffic, and responses to stress, including the immune response. In many of these settings, Ub acts through routes that involve the processive degradation of Ub-protein conjugates (Hershko 1991; Jentsch 1992; Varshavsky 1996a).

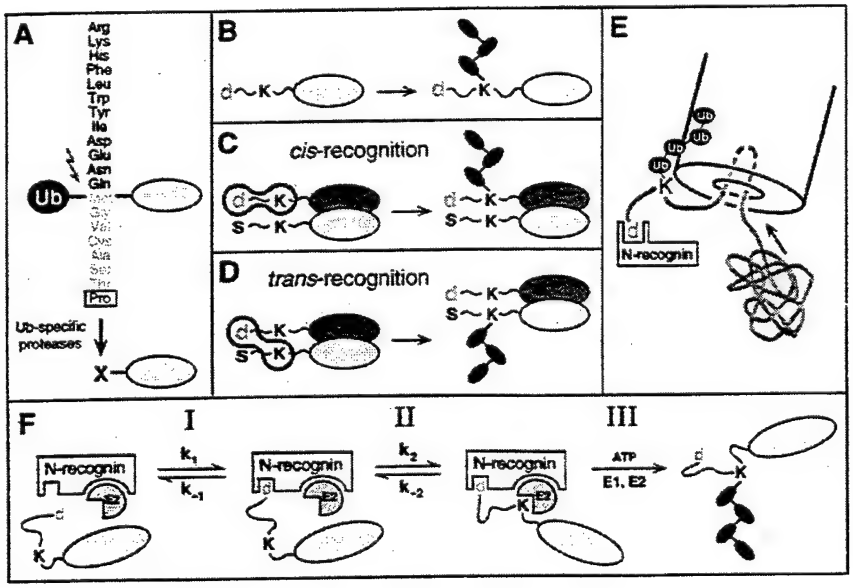


**Table 1** The N-end rule in *E. coli* and *S. cerevisiae*

| Residue X | Half-life of X-βgal | |
|---|---|---|
| | *E. coli* | *S. cerevisiae* |
| Arg | 2 min | 2 min |
| Lys | 2 min | 3 min |
| Phe | 2 min | 3 min |
| Leu | 2 min | 3 min |
| Trp | 2 min | 3 min |
| Tyr | 2 min | 10 min |
| His | > 10 h | 3 min |
| Ile | > 10 h | 30 min |
| Asp | > 10 h | 3 min |
| Glu | > 10 h | 30 min |
| Asn | > 10 h | 3min |
| Gln | > 10 h | 10 min |
| Cys | > 10 h | > 30 h |
| Ala | > 10 h | > 30 h |
| Ser | > 10 h | > 30 h |
| Thr | > 10 h | > 30 h |
| Gly | > 10 h | > 30 h |
| Val | > 10 h | > 30 h |
| Pro | ? | > 5 h |
| Met | > 10 h | > 30 h |

Approximate *in vivo* half-lives of X-βgal proteins in *E. coli* at 36 °C (Tobias *et al.* 1991) and in *S. cerevisiae* at 30 °C (Bachmair & Varshavsky 1989). A question mark at Pro indicates its uncertain status in the N-end rule (see the main text).

The binding of an N-end rule substrate by a targeting complex is followed by the formation of a substrate-linked multi-Ub chain (Dohmen *et al.* 1991). The ubiquitylated substrate is processively degraded by the 26S proteasome—an ATP-dependent, multisubunit protease (Rechsteiner *et al.* 1993). The N-end rule pathway is present in both the cytosol (Bachmair *et al.* 1986) and the nucleus (J.A. Johnston & A.V., unpublished data). In this paper, I summarize the current understanding of the N-end rule. For a more detailed review, see Varshavsky *et al.* (1997).

## Components and evolution of the N-end rule pathway

The N-end rule is organized hierarchically. In *S. cerevisiae*, Asn and Gln are *tertiary* destabilizing N-terminal residues in that they function through their conversion, by enzymatic deamidation, into the *secondary* destabilizing N-terminal residues Asp and Glu, whose activity requires their conjugation, by Arg-tRNA-protein transferase (R-transferase), to Arg, one of the *primary* destabilizing N-terminal residues (Gonda *et al.* 1989; Balzi *et al.* 1990; Baker & Varshavsky 1995). The primary destabilizing residues are bound directly by N-recognin (also called E3), the recognition component of the N-end rule pathway (Fig. 3).

### N-recognin (E3)

In *S. cerevisiae*, N-recognin is a 225 kDa protein (encoded by *UBR1*) that selects potential N-end rule

**Figure 2** Mechanics of the N-end rule. (**A**) The ubiquitin fusion technique. Linear fusions of Ub to other proteins are cleaved at the last residue of Ub, making it possible to produce different residues at the N-termini of otherwise identical proteins (Bachmair *et al.* 1986; Gonda *et al.* 1989). Amino acid residues in blue and red are stabilizing and destabilizing, respectively, in the *S. cerevisiae* N-end rule (Bachmair & Varshavsky 1989). (**B**) The two-determinant organization of eukaryotic N-degrons. *d*, a destabilizing N-terminal residue. A chain of black ovals linked to the second-determinant lysine (*K*) denotes a multi-Ub chain. (**C**) *Cis* recognition of the N-degron in one subunit of a dimeric protein. The other subunit bears *s*, a stabilizing N-terminal residue. (**D**) *Trans* recognition, in which the first (*d*) and second (*K*) determinants of the N-degron reside in different subunits of a dimeric protein (Johnson *et al.* 1990). (**E**) The hairpin insertion model. A targeted N-end rule substrate (in green) bearing a multi-Ub chain is shown bound to the 26S proteasome through the chain. The position of a targeting complex containing N-recognin is unknown, and is left unspecified. Only the 20S core component of the 26S proteasome is shown. A red arrow indicates the direction of net movement of the substrate's polypeptide chain toward active sites in the interior of proteasome. By analogy with the arrangement of signal sequences during transmembrane translocation of proteins (Schatz & Dobberstein 1996), it is proposed that a region of the substrate upstream of its ubiquitylated lysine (K) does not move through the proteasome during the substrate's degradation, and may be released intact following a cleavage downstream of the lysine. Variants of this model may also be relevant to the targeting of proteins that bear internal or C-terminal degrons. (**F**) A model for the recognition of an N-end rule substrate (Bachmair & Varshavsky 1989). The reversible binding of N-recognin to a primary destabilizing N-terminal residue (*d*) of a substrate (step **I**) must be followed by a capture of the second-determinant lysine (*K*) of the substrate by a targeting complex containing a Ub-conjugating (E2) enzyme (step **II**). It is unknown whether the lysine is captured by E2 (as shown here) or by N-recognin. Ubiquitylation of the substrate commences once the targeting complex is bound to both determinants of the N-degron (step **III**). This model does not specify, among other things, the details of Ub conjugation (see the main text).

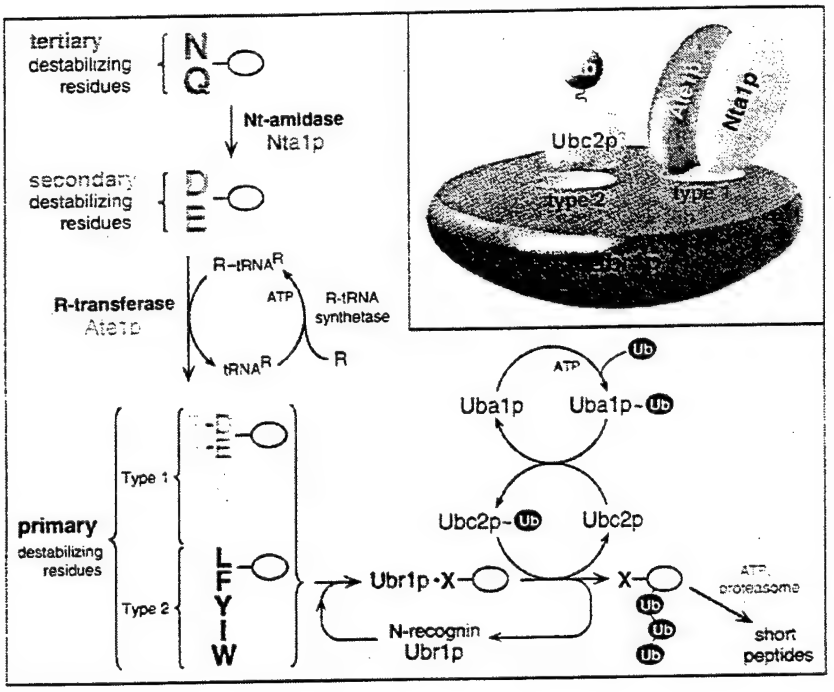



**Figure 3** The *S. cerevisiae* N-end rule pathway. Type 1 and type 2 primary destabilizing N-terminal residues are in purple and red, respectively. Secondary and tertiary destabilizing N-terminal residues are in blue and green, respectively. The yellow ovals denote the rest of a protein substrate. The conversion of tertiary destabilizing residues **N** and **Q** into secondary destabilizing residues **D and E** is mediated by N-terminal amidohydrolase (Nt-amidase), encoded by *NTA1*. The conjugation of the primary destabilizing residue **R** to the secondary destabilizing residues **D** and **E** is mediated by Arg-tRNA-protein transferase (R-transferase), encoded by *ATE1*. A complex of N-recognin and the Ub-conjugating (E2) enzyme Ubc2p catalyses the conjugation of activated Ub, produced by the Ub-activating (E1) enzyme Uba1p, to a Lys residue of the substrate, yielding a substrate-linked multi-Ub chain. Uba1p~Ub and Ubc2p~Ub denote covalent (thioester-mediated) complexes of these enzymes with Ub. A multiubiquitylated substrate is degraded by the 26S proteasome. Inset: A model of the targeting complex. The 20 kDa Ubc2p E2 enzyme is depicted carrying activated Ub linked to Cys-88 of Ubc2p through a thioester bond. Both the 52 kDa Nta1p (Nt-amidase) and the 58 kDa Ate1p (R-transferase) bind to the 225 kDa Ubr1p (N-recognin) in proximity to the type 1 substrate-binding site of Ubr1p. In addition, Nta1p directly interacts with Ate1p (see the main text).

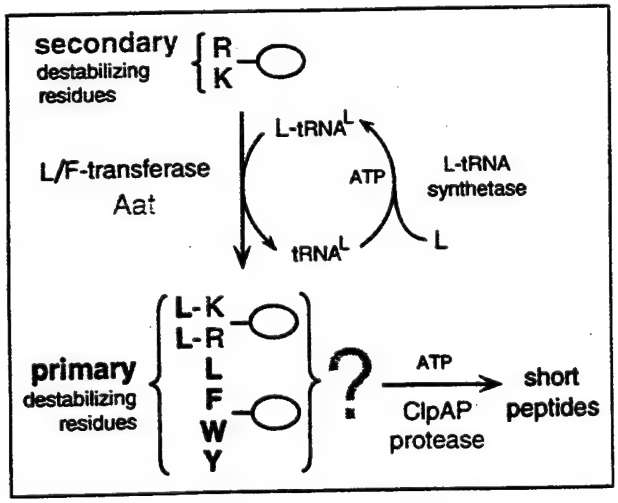



**Figure 4** The *E. coli* N-end rule pathway. Primary destabilizing N-terminal residues **L**, **F**, **W** and **Y** are in red. Secondary destabilizing N-terminal residues **R** and **K** are in blue. The yellow ovals denote the rest of a protein substrate. Conjugation of the primary destabilizing residue **L** to the secondary destabilizing residues **R** and **K** is mediated by Leu, Phe-tRNA-protein transferase (L/F-transferase), encoded by *aat* (Tobias *et al.* 1991). *In vivo*, L/F-transferase appears to conjugate predominantly, if not exclusively, **L** (Shrader *et al.* 1993). The degradation of a substrate bearing a primary destabilizing N-terminal residue is carried out by the ATP-dependent protease ClpAP, encoded by *clpA* and *clpP*. A question mark denotes an ambiguity about the nature of N-recognin in *E. coli*.

substrates through the binding to their primary destabilizing N-terminal residues Phe, Leu, Trp, Tyr, Ile, Arg, Lys or His (Bartel *et al.* 1990; Varshavsky 1996a). N-recognin has at least two substrate-binding sites. The type 1 site is specific for the basic N-terminal residues Arg, Lys and His. The type 2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr and Ile (Fig. 3). At present, these sites are defined through dipeptide-based competition experiments. Specifically, a dipeptide bearing a destabilizing N-terminal residue was found to inhibit the degradation of a test N-end rule substrate if that substrate's N-terminal residue was of the same type (1 or 2) as the dipeptide's N-terminal residue (Reiss *et al.* 1988; Gonda *et al.* 1989).

A genetic dissection of the type 1 and type 2 sites in *S. cerevisiae* N-recognin (Ubr1p) has shown that either of the sites can be mutationally inactivated without significantly perturbing the other site. Mutations that selectively inactivate the type 1 or the type 2 site are located within the ≈50 kDa N-terminal region of the 225 kDa N-recognin (A. Webster, M. Ghislain & A. V., unpublished data). E3$\alpha$, the mammalian counterpart of *S. cerevisiae* N-recognin, has been characterized biochemically in extracts from rabbit reticulocytes (Hershko 1991). Another mammalian N-recognin, termed E3$\beta$, which apparently binds to substrates bearing N-terminal Ala and Thr (and possibly also Ser (Lévy *et al.* 1996)), has been described as well (Hershko 1991).

All eukaryotes examined have both Ub and the N-end rule pathway. Some, but not all, prokaryotes contain Ub (Wolf *et al.* 1993). The bacterium *E. coli* lacks Ub but does have an N-end rule pathway (Fig. 4) (Tobias *et al.* 1991). Screens for mutations that inactivate either the entire N-end rule pathway or its subset have identified three *E. coli* genes—*clpA*, *clpP*, and *aat* (Shrader *et al.* 1993). Aat is a Leu, Phe-tRNA-protein transferase (L/F-transferase). ClpA (81 kDa) and ClpP (21 kDa) form a ≈750 kDa complex, ClpAP, which exhibits ATP-dependent protease activity *in vitro* (Gottesman & Maurizi 1992), and is a functional counterpart of the eukaryotic 26S proteasome in the *E. coli* N-end rule pathway (Fig. 4).

ClpP exhibits a chymotrypsin-like protease activity *in vitro* (Gottesman & Maurizi 1992). ClpA is the ATP-binding component of ClpAP. *In vitro* studies have shown that ClpA can act as a chaperone in the activation of RepA, the replication initiator encoded by the plasmid P1 (Wickner *et al.* 1994). *In vivo* ramifications of these results, and in particular their relevance to the proteolytic function of ClpAP in the *E. coli* N-end rule pathway (Fig. 4), remain to be examined. ClpP associates not only with ClpA (forming ClpAP protease), but also with the ClpA homologues ClpB or ClpX, forming, respectively, ClpBP or ClpXP proteases (Gottesman *et al.* 1993; Wawrzynov *et al.* 1995). In contrast to ClpA, whose mutational elimination stabilizes the normally short-lived N-end rule substrates (Tobias *et al.* 1991), the elimination of either ClpB or ClpX appears not to perturb the *E. coli* N-end rule pathway (O. Lomovskaya & A. V., unpublished data).

## N-terminal amidases

The *S. cerevisiae* N-terminal amidohydrolase (Nt-amidase), encoded by *NTA1*, is a 52 kDa enzyme which deamidates Asn or Gln if, and only if, they are located at the N-terminus of a polypeptide (Figs 3 and 5B) (Baker & Varshavsky 1995; Grigoryev *et al.* 1996). Null *nta1* mutants are unable to degrade N-end rule substrates that bear N-terminal Asn or Gln. The deduced sequence of Nta1p is not similar to those of the other known amidotransferases, save for the sequence Gly Ile-Cys-Met that is a part of an 11-residue region conserved among some, but not all, amidotransferases. The conserved cysteine of this sequence is required for the enzymatic activity of Nta1p (Grigoryev *et al.* 1996).

Stewart *et al.* (1995) purified a porcine Nt-amidase that deamidates N-terminal Asn (N) but not Gln (Q), and isolated a cDNA that encodes this enzyme. Grigoryev *et al.* (1996) isolated and characterized a ≈17 kb gene, termed *Ntan1*, that encodes a mouse homologue of the porcine amidase, termed $Nt^N$-amidase. The ≈1.4 kb *Ntan1* mRNA is expressed in all of the tested mouse tissues and cell lines. The recently produced *ntan1*$\Delta$ mouse strains are viable (Y.T. Kwon and A.V., unpublished data). Their phenotypic analysis is under way.

Both Asn and Gln are destabilizing residues in the mammalian N-end rule (Fig. 1). Further, both N-terminal Asn and Gln of the test proteins are deamidated in mammalian cell extracts (Gonda *et al.* 1989; S. Grigoryev & A.V., unpublished data). Therefore there must exist yet another mammalian Nt-amidase ($Nt^Q$-amidase), which can deamidate N-terminal Gln (Fig. 5A).

## Aminoacyl-tRNA-protein transferases

The *S. cerevisiae* Arg-tRNA-protein transferase (R-transferase), encoded by *ATE1*, is a 58 kDa enzyme that utilizes Arg-tRNA to arginylate the N-termini of polypeptides (but not free amino acids) that bear Asp or Glu (Figs 3 and 5B). Null *ate1* mutants are unable to

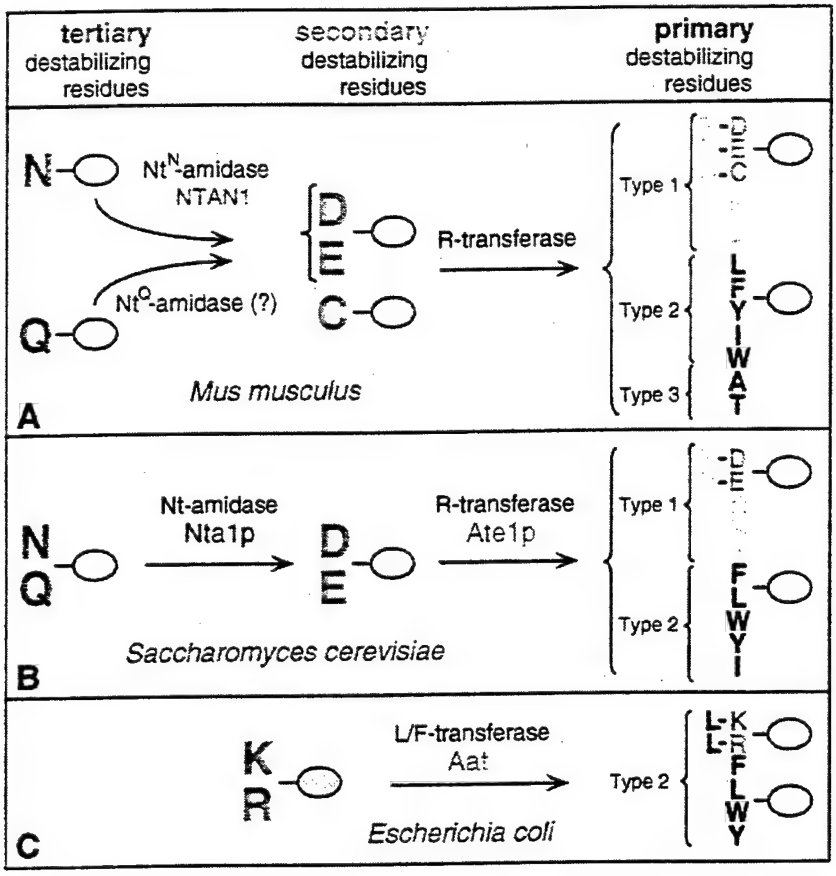



**Figure 5** Comparison of enzymatic reactions that underlie the activity of tertiary and secondary destabilizing residues in different organisms. (**A**) Mouse (*Mus musculus*) L-cells and rabbit (*Oryctolagus cuniculus*) reticulocytes (Lévy *et al.* 1996; Gonda *et al.* 1989). (**B**) The yeast *Saccharomyces cerevisiae* (Bachmair & Varshavsky 1989). (**C**) The bacterium *Escherichia coli* (Tobias *et al.* 1991). The *E. coli* N-end rule lacks tertiary destabilizing residues. The postulated mammalian $Nt^Q$-amidase (a question mark in **A**) remains to be identified.

degrade N-end rule substrates that bear N-terminal Asn, Gln, Asp or Glu—the tertiary and secondary destabilizing N-terminal residues (Balzi *et al.* 1990). In contrast to *S. cerevisiae*, where only Asp and Glu are secondary destabilizing residues, in mammals, Cys is a secondary destabilizing residue as well (Gonda *et al.* 1989) (Table 1 and Fig. 1). It is not known whether the arginylation of N-terminal Asp, Glu and Cys in mammals is catalysed by an R-transferase whose specificity is broader than that of its yeast counterpart, or whether N-terminal Cys is arginylated by a distinct R-transferase (Fig. 5).

Approximately 2 h after a crush injury to the rat sciatic nerve, an extract was prepared from a segment of the nerve immediately upstream of the crush site. This extract was found to conjugate a ≈10-fold higher amount of the added $^3$H-arginine to the N-termini of unidentified endogenous proteins than an otherwise identical extract from the same region of an unperturbed sciatic nerve (Dayal *et al.* 1990), suggesting a crush-induced increase in the level of N-end rule substrates and/or a post-crush induction of the N-end rule pathway. No post-crush increase in arginylation was observed with extracts from the optic nerve, which does not regenerate after a crush injury, in contrast to the sciatic nerve (Dayal *et al.* 1990).

R-transferase appears to be confined to eukaryotes, whereas Leu, Phe-tRNA-protein transferase (L/F-transferase) is present in bacteria such as *E. coli* but is apparently absent from eukaryotes. *E. coli* L/F-transferase is a 27 kDa enzyme encoded by the gene *aat* (Shrader *et al.* 1993). *In vivo*, L/F-transferase conjugates mainly if not exclusively the Leu to N-terminal Arg or Lys of a polypeptide substrate (Shrader *et al.* 1993) (Fig. 4). *E. coli* mutants lacking *aat* are unable to degrade N-end rule substrates that bear N-terminal Arg or Lys. These data (Tobias *et al.* 1991) identified L/F-transferase as a component of the *E. coli* N-end rule pathway.

## Ubiquitin-conjugating enzymes

The initial interaction between an N-end rule substrate and N-recognin is of moderate affinity (the inferred $K_d$ of roughly 10 μM; Varshavsky *et al.* 1997), but becomes much stronger if an internal lysine of the

substrate is captured by a targeting complex containing a Ub-conjugating (E2) enzyme and N-recognin (E3). This capture initiates a processive synthesis of a lysine-linked multi-Ub chain. The E2 enzymes utilize activated Ub, produced by the Ub-activating (E1) enzyme, to catalyse the formation of isopeptide bonds between the C-terminal Gly 76 of Ub and ε-amino groups of lysines in acceptor proteins (Fig. 3) (Pickart 1988; Hershko 1991; Jentsch 1992).

In at least some Ub-dependent systems (Scheffner *et al.* 1995), including apparently the N-end rule pathway (V. Chau & A.V., unpublished data), the pathway-specific Ub ligase—a complex of a recognin (E3) and an E2 enzyme—shifts the activated Ub moiety (which is initially linked to a Cys residue of the E1 enzyme) through a relay of Ub thioesters before conjugating Ub to a Lys residue of a targeted substrate. In a substrate-linked multi-Ub chain, the C-terminal glycine of one Ub moiety is joined to an internal lysine of the adjacent Ub moiety, resulting in a chain of Ub-Ub conjugates. In a multi-Ub chain linked to an N-end rule substrate, only Lys-48 of Ub was found to be joined to another Ub moiety within a chain (Chau *et al.* 1989). Recently, multi-Ub chains mediated by Lys-63, Lys-29, Lys-11 or Lys-6 of Ub have been described as well (Arnason & Ellison 1994; Johnson *et al.* 1995; Spence *et al.* 1995; Baboshina & Haas 1996). It is not known whether these chains play a role in the N-end rule pathway.

In *S. cerevisiae*, the ubiquitylation of N-end rule substrates requires the Ubc2p E2 enzyme (Dohmen *et al.* 1991). Processes known to be perturbed by mutations in *UBC2* include the N-end rule pathway, DNA repair, induced mutagenesis, cell cycle control, and sporulation (Jentsch 1992, and references therein). The N-end rule pathway is inactive in both *ubr1*Δ and *ubc2*Δ mutants (Madura *et al.* 1993), but the overall effect of *ubc2*Δ on cell growth and sporulation is much more severe than that of *ubr1*Δ, indicating that the functions of Ubc2p are not confined to the N-end rule pathway.

## The N-end rule as a witness of evolution

The organization of N-end rules, with their tertiary, secondary and primary destabilizing residues, is a feature more conserved in evolution than either the Ub dependence of an N-end rule pathway or the identity of enzymatic reactions that mediate the hierarchy of destabilizing residues. For example, in a bacterium such as *E. coli*, which lacks the Ub system, the N-end rule has both secondary and primary destabilizing residues (it lacks tertiary residues) (Figs 1, 4 and 5C). The identities of secondary destabilizing residues in *E. coli* (Arg and Lys) are different from those in eukaryotes (Figs 1 and 5). Bacterial and eukaryotic enzymes that implement the coupling between secondary and primary residues are also different: L/F-transferase in *E. coli* and R-transferase in eukaryotes. Note, however, that bacterial L/F-transferase and eukaryotic R-transferase catalyse reactions of the same type (conjugation of an amino acid to an N-terminal residue of a polypeptide) and utilize the same source of activated amino acid (aminoacyl-tRNA) (Fig. 5).

The apparent confinement of R-transferase to eukaryotes and of L/F-transferase to prokaryotes suggests that secondary destabilizing residues were recruited late in the evolution of N-end rule, after the divergence of prokaryotic and eukaryotic lineages. The lack of sequence similarity between the yeast Nt-amidase and the mammalian $Nt^N$-amidase, as well as the more narrow specificity of the mammalian enzyme (Fig. 5A, B) suggest that tertiary destabilizing residues Asn and Gln became a part of the N-end rule much later yet, possibly after the divergence of metazoan and fungal lineages. If so, the N-end rule pathway may be an especially informative witness of evolution: the ancient origins of this proteolytic system, the simplicity and discreteness of changes in the rule books of N-end rules among different species, and the diversity of proteins that either produce or target the N-degron should facilitate phylogenetic deductions—once the components of this pathway become characterized across a broad range of organisms.

## Code vs. hardware

A given N-end rule is defined operationally—for a set of proteins such as X-βgals that differ exclusively by their N-terminal residues. Existing evidence (Bachmair & Varshavsky 1989) suggests that the ranking aspect of an N-end rule, i.e., an ordering of relative destabilizing activities among 20 fundamental amino acids, is invariant from one protein reporter to another in a given intracellular compartment. (The case of N-end rule substrates bearing N-terminal Pro presents an apparent exception to this conjecture; see below.) By contrast, the actual *in vivo* half-lives may differ greatly among *different* proteins bearing one and the same N-terminal residue (Bachmair & Varshavsky 1989). The cause of these differences is the multicomponent nature of underlying N-degrons (Fig. 2B). For example, in eukaryotes, an N-degron comprises not only a destabilizing N-terminal residue of a

protein but also its internal lysine (or lysines), whose quality as a determinant can range from high to nonexistent.

*A priori*, one and the same N-end rule can be implemented through vastly different assortments of targeting hardware. At one extreme, each destabilizing N-terminal residue may be bound by a distinct N-recognin. Conversely, a single N-recognin may be responsible for the entire rule book of destabilizing residues in a given N-end rule. The actual N-end rule pathways lie between these extremes, and happen to have a hierarchic rather than 'linear' structure (Figs 3–5).

## Targeting complex of the N-end rule pathway

The known components of the *S. cerevisiae* N-end rule pathway that mediate steps prior to the proteolysis of a targeted substrate by the 26S proteasome are Nt-amidase (Nta1p), R-transferase (Ate1p), N-recognin (Ubr1p), a Ub-conjugating (E2) enzyme (Ubc2p), and the Ub-activating (E1) enzyme (Uba1p) (Fig. 3) (Varshavsky 1996a). In addition to direct (immunoprecipitation-based) evidence for the physical association between N-recognin and Ubc2p (Madura *et al.* 1993), there is also circumstantial (overexpression-based) evidence for the existence of a complex between N-recognin, R-transferase and Nt-amidase (Baker & Varshavsky 1995). Recently, a high–affinity interaction between Nta1p and Ate1p was demonstrated directly; other data suggest that both Nta1p and Ate1p interact with Ubr1p (M. Ghislain, A. Webster & A.V., unpublished results). In a quaternary Ubc2p-Ubr1p-Nta1p-Ate1p complex suggested by these data, Ate1p and Nta1p interact with each other and Ubr1p (Fig. 3).

Other, perhaps more transient, components of the targeting complex in *S. cerevisiae* are likely to include the 114 kDa Uba1p (E1 enzyme, which must be bound to Ubc2p during the E1→E2 transfer of activated Ub moiety), and also Arg-tRNA synthetase. The latter possibility is suggested by the finding that, in mammals, Arg-tRNA synthetase (whose product, Arg-tRNA, is a co-substrate of R-transferase) copurifies with R-transferase (Ciechanover *et al.* 1988). It is also likely that the targeting complex interacts with the 26S proteasome *in vivo*, for example, during the transfer of a multiubiquitylated N-end rule substrate to substrate-binding sites of the proteasome. The proteolytic machine that implements the N-end rule is thus a strikingly diverse assembly of enzymes and binding factors whose total mass is close to that of the large ribosomal subunit. However, even a transient existence of this 'metacomplex' is conjectural at present, the alternative possibility being a sequential formation of transient subcomplexes that produce a substrate-linked multi-Ub chain and relay a substrate toward the 26S proteasome.

The effects of overexpressing Nt-amidase and/or R-transferase in *S. cerevisiae* not only suggested the existence of the Nta1p-Ate1p-Ubr1p-Ubc2p complex but also led to the prediction that Nta1p and Ate1p are associated with Ubr1p in proximity to its type 1 substrate-binding site (Fig. 3) (Baker & Varshavsky 1995). The 'proximity' aspect of the postulated complex was invoked to account for the markedly different effects of overexpressed R-transferase on the degradation of N-end rule substrates bearing type 1 vs. type 2 primary destabilizing N-terminal residues (Baker & Varshavsky 1995). In the diagram of Fig. 3, the physical proximity of the bound R-transferase to the type 1 site of N-recognin is presumed to decrease the steric accessibility of this site to an N-end rule substrate that bears a type 1 residue such as Arg and approaches the type 1 binding site of N-recognin directly from the bulk solvent. By contrast, a substrate that acquired Arg through arginylation by the N-recognin-bound R-transferase would be able to reach the (nearby) type 1 binding site of N-recognin directly—without dissociating into the bulk solvent first—a feature known as substrate 'channeling' in multistage enzymatic reactions (Negrutskii & Deutscher 1991). The mechanics of channeling may involve the diffusion of an N-end rule substrate in proximity to surfaces of the targeting complex, analogous to the mechanism of a bifunctional enzyme dihydrofolate reductase-thymidylate synthetase, where the channeling of dihydrofolate apparently results from its movement across the surface of the protein (Knighton *et al.* 1994).

## The N-degron and pre-N-degron

Nascent proteins contain N-terminal Met (fMet in prokaryotes), which is a stabilizing residue in the known N-end rules (Fig. 1). Thus, the N-degron of an N-end rule substrate must be produced from a pre-N-degron. In an engineered N-end rule substrate, a pre-N-degron contains the N-terminal Ub moiety whose removal by Ub-specific proteases yields the protein's N-degron (Fig. 2A). This design of a pre-N-degron is unlikely to be relevant to physiological N-end rule substrates, because natural Ub fusions (including the precursors of Ub) either contain a stabilizing residue at the Ub-protein junction or bear a mutant Ub moiety that is

retained *in vivo* (Özkaynak *et al.* 1987; Finley *et al.* 1989; Watkins *et al.* 1993). The known Met-aminopeptidases remove N-terminal Met if, and only if, the second residue of a protein is stabilizing in the yeast N-end rule (Fig. 1). The structural basis of this selectivity is the size of a residue's side chain (Sherman *et al.* 1985; Arfin & Bradshaw 1988; Li & Chang 1995). Specifically, the side chains of the residues that are destabilizing in the yeast N-end rule are larger than those of stabilizing residues. The exception is Met—a bulky hydrophobic but stabilizing residue (Fig. 1).

Can there be just one or a few residues between N-terminal Met and the site of cleavage that produces an N-degron? If so, a short (≤10 residues) N-terminal sequence might contain both the recognition motif and the cleavage site(s) for a relevant (unknown) processing protease. Screens for such sequences, carried out in *S. cerevisiae* (Sadis *et al.* 1995; Ghislain *et al.* 1996), did identify short (≤10 residues) N-terminal regions that conferred Ubr1p-dependent metabolic instability on a reporter protein. Most of the sequences identified by these screens were not similar to each other, possibly because a very large number of 10-residue N-terminal extensions can produce an N-degron *in vivo*, analogous to a large number of N-terminal sequences that can function as signals for protein translocation across the ER membrane (Kaiser *et al.* 1987).

Analysis of one N-terminal extension identified by Ghislain *et al.* (1996) has shown that it targets a reporter protein for degradation while retaining its N-terminal Met (M. Gonzalez, F. Lévy, M. Ghislain & A.V., unpublished data). This finding suggests that N-recognin binds not only to N-degrons but also to a degron that consists of an entirely internal sequence motif. By contrast, two other examined (directly sequenced) extensions were found to be cleaved *in vivo* after N-terminal Met, yielding destabilizing N-terminal residues (Sadis *et al.* 1995; Ghislain *et al.* 1996). In sum, we are just beginning to understand the processing reactions that yield a destabilizing N-terminal residue in a non-polyprotein context.

## Mechanics of N-degron

### Stochastic capture model

Studies with βgal- and DHFR-based N-end rule substrates (Bachmair & Varshavsky 1989; Chau *et al.* 1989; Johnson *et al.* 1990; Hill *et al.* 1993) suggested a stochastic view of the N-degron, in which specific lysines of an N-end rule substrate can be assigned a probability of being utilized as a ubiquitylation site. This probability depends on the time-averaged spatial position and mobility of a protein's lysine. For some, and often for most of the lysines in an N-end rule substrate, the probability of serving as a ubiquitylation site would be negligible because of the lysine's lack of mobility and/or its distance from a destabilizing N-terminal residue. In this 'stochastic capture' model (Fig. 2F), the folded conformation of a substrate would be expected to slow down or preclude the search for a Lys residue, unless it is optimally positioned in the folded substrate.

The bipartite design of N-degron (Fig. 2B) is also likely to be characteristic of other Ub-dependent degradation signals—present in a multitude of naturally short-lived proteins that include cyclins (Murray & Hunt 1993), IκBα, and c-Jun (Pahl & Baeuerle 1996). The first component of these degrons is the internal region of a protein (instead of its N-terminal residue) that is specific for each degradation signal. The second component is an internal lysine (or lysines). A degron may also contain regulatory determinants whose modification (e.g. phosphorylation/dephosphorylation) can modulate the activity of this degron (Pahl & Baeuerle 1996; Nishizawa *et al.* 1993).

### *Cis-trans* recognition and subunit-specific degradation of oligomeric proteins

The two determinants of N-degron can be recognized either in *cis* or in *trans* (Fig. 2C, D) (Johnson *et al.* 1990; F. Lévy & A. V., unpublished data). Experiments that revealed the *trans*-recognition have also brought to light a remarkable feature of the N-end rule pathway: only those subunits of an oligomeric protein that contain the ubiquitylation site (but not necessarily a destabilizing N-terminal residue) are actually degraded (11).

What might be the mechanism of subunit-specific proteolysis? A 'simple' model is suggested by the binding of a substrate-linked multi-Ub chain to a component of the proteasome. Specifically, a subunit of an oligomeric substrate bound to the proteasome through a subunit-linked multi-Ub chain may be the only subunit that undergoes further mechanochemical processing by chaperone-like, ATP-dependent components of the 26S proteasome. These components mediate the unfolding and translocation steps that cause a movement of the subunit toward active sites in the proteasome's interior, and in the process dissociate this subunit from the rest of oligomeric substrate. In this mechanism, the initial binding of N-recognin to another subunit—that which bears a destabilizing

N-terminal residue but not the lysine determinant (Fig. 2C)—may be either too transient (lasting, in a 'productive' engagement, only long enough for a lysine to be captured on a nearby subunit) or sterically unfavourable for the delivery of this subunit to the interior of the proteasome.

Since other Ub-dependent degradation signals appear to be organized similarly to the N-degron (a 'primary' recognition determinant plus an internal lysine or lysines), subunit selectivity is likely to be a general feature of proteolysis by the Ub system (Varshavsky 1996a). Examples of physiologically relevant subunit-selective proteolysis include the degradation of p53 in a complex with the papilloma viral protein E6 (Scheffner *et al.* 1995) and the degradation of a cyclin in a complex with a cyclin-dependent kinase (Murray & Hunt 1993).

## The hairpin insertion model and the function of multiubiquitin chain

Formation of a substrate-linked multi-Ub chain produces an additional binding site (or sites) for components of the proteasome. The resulting increase in affinity, i.e. a decrease in the rate of dissociation of the proteasome–substrate complex, can be used to facilitate proteolysis. Suppose that a rate-limiting step which leads, several steps later, to the first proteolytic cleavage of the proteasome-bound substrate is an unfolding (driven by thermal fluctuations) of a relevant region of the substrate. If so, an increase in stability of the proteasome–substrate complex, brought about by the multi-Ub chain, should facilitate the substrate's degradation, because the longer the allowed 'waiting' time, the greater the probability of a required unfolding event. Another (not mutually exclusive) possibility is that a substrate-linked multi-Ub chain acts as a proximity trap for partially unfolded states of a substrate. This might be achieved through reversible interactions of the chain's Ub moieties with regions of the substrate that undergo local unfolding. A prediction common to both models is that the degradation of a substrate whose conformation poses less of a kinetic impediment to the proteasome should be less dependent on Ub and ubiquitylation than the degradation of an otherwise similar but more stably folded substrate.

How is a proteasome-bound, ubiquitylated protein directed to the interior of the proteasome? This problem is analogous to that in studies of transmembrane channels for protein translocation (Simon & Blobel 1991; Schatz & Dobberstein 1996). Could the solutions be similar in these systems, reflecting, perhaps, a common ancestry of translocation channels and proteasomes? The model in Fig. 2E proposes, by analogy with translocation systems, a 'hairpin' insertion mechanism for the initiation of proteolysis by the 26S proteasome. A biased random walk ('thermal ratchet') that is likely to underlie the translocation of proteins across membranes (Simon & Blobel 1991) may also be responsible for the movement of the substrate's polypeptide chain through the proteasome, with cleavage products diffusing out from the proteasome's distal end and thereby contributing to the net bias in the chain's bidirectional saltations through the proteasome channel. One prediction of the hairpin insertion model for an N-end rule substrate whose N-degron's determinants are located upstream of the hairpin insertion site is that the substrate's N-terminal region (Fig. 2E) is likely be cleaved-off at later stages of targeting, and is therefore likely to be spared from the proteasome-mediated degradation.

Two findings indicate that the unfolding of a targeted N-end rule substrate is a prerequisite for its degradation by the 26S proteasome. Methotrexate—a folic acid analogue and high-affinity ligand of DHFR—can inhibit the degradation of an N-end rule substrate such as Arg-DHFR by the N-end rule pathway (Johnston *et al.* 1995). This result suggests that a critical post-ubiquitylation step faced by the proteasome includes a 'sufficient' conformational perturbation of the proteasome-bound substrate. Furthermore, it was shown that the N-end rule-mediated degradation of a 17 kDa N-terminal fragment of the 70 kDa Sindbis virus polymerase is not precluded by the conversion of all of the fragment's 10 Lys residues into Arg residues, which cannot be ubiquitylated (T. Rümenapf, J. Strauss & A.V., unpublished data). Thus, the ubiquitylation requirement of previously studied N-end rule substrates may be a consequence of their relatively stable conformations. The binding of a largely unfolded substrate (such as a fragment of Sindbis polymerase) by the targeting complex of the N-end rule pathway may be sufficient for the delivery of the substrate to the proteasome's active sites in the absence of a multi-Ub chain. In the language of models in Fig. 2E, F, the 'waiting' time for a bound and conformationally unstable substrate may be short enough not to require the formation of a dissociation-slowing device such as a multi-Ub chain.

## The N-end rule without ubiquitin

No Ub-like covalent modification of N-end rule substrates has been detected in *E. coli*, in contrast to ubiquitylation of the same substrates in eukaryotes.

Moreover, the conversion of ubiquitylation-site lysines of an N-end rule substrate into arginines rendered the substrate long-lived in eukaryotes but did not impair its degradation in *E. coli* (Tobias *et al.* 1991). Thus, *E. coli* not only lacks a homologue of eukaryotic Ub, but also lacks the requirement for a lysine-specific modification of a substrate. Bacteria may contain proteins whose function in the N-end rule pathway is Ub-like but involves a noncovalent, lysine-independent binding to a targeted substrate. The proposed role of a substrate-linked multi-Ub chain in 'marking' a subunit of a protein for selective destruction leads to another testable conjecture: if a subunit-marking device is absent from the *E. coli* N-end rule pathway, the latter may be incapable of degrading an oligomeric protein 'one subunit at a time'.

## Substrates and functions of the N-end rule pathway

### The N-end rule and osmoregulation in yeast

A synthetic lethal screen was used to isolate an *S. cerevisiae* mutant, termed *sln1* (for 'synthetic lethal of N-end rule'), whose viability requires the presence of *UBR1* (Ota & Varshavsky 1992). *SLN1* has been found to encode a eukaryotic homologue of two-component regulators—a large family of proteins previously encountered only in bacteria (Ota & Varshavsky 1993). The properties of *S. cerevisiae* Sln1p are consistent with it being a sensor component of the osmoregulatory (HOG) pathway—a MAP kinase cascade (Maeda *et al.* 1994). Since an otherwise lethal hypomorphic mutation in *SLN1* can be suppressed by the presence of Ubr1p (N-recognin) (Ota & Varshavsky 1993), it is likely that one or more of the proteins (e.g., kinases) whose activity is down-regulated by Sln1p can also be down-regulated through their degradation by the N-end rule pathway. The relevant physiological N-end rule substrate(s) remains to be identified.

### The N-end rule and the import of peptides

Alagramam *et al.* (1995) have found that *ubr1*Δ yeast cells are unable to import di- and tripeptides. They have also shown that *ubr1*Δ cells, unlike the congenic *UBR1* cells, contain virtually no *PTR2* mRNA that encodes a peptide transporter, an integral plasma membrane protein. Recent results (C. Byrd & A.V., unpublished data) indicated that the control of *PTR2* expression by Ubr1p (N-recognin) involves the Ub-conjugating (E2) enzyme Ubc2p, a known component of the N-end rule pathway (Fig. 3). The Ubc4p E2 enzyme can partially compensate for the absence of Ubc2p; deletion of both *UBC2* and *UBC4* results in cells that do not express Ptr2p and are unable to import peptides, similarly to *ubr1*Δ cells. Ubc4p has not been previously identified as a component of the N-end rule pathway.

The findings of Alagramam *et al.* (1995), and the observed dependence of peptide import on two specific ubiquitin-conjugating enzymes can be accounted for by a model in which the expression of the Ptr2p transporter is regulated by a short-lived repressor that is degraded by the N-end rule pathway. One prediction of this model is that a mutational inactivation of the repressor would bypass the requirement for Ubr1p in the import of peptides. Using a screen for such mutants, we isolated a gene called *CUP9* (C. Byrd & A.V., unpublished data). Its product is a homeodomain-containing, short-lived protein whose degradation is carried out largely by the N-end rule pathway. Overexpression of Cup9p inhibits the import of peptides. Conversely, *cup9*Δ cells express Ptr2p and import peptides at higher rates than *CUP9* cells. Moreover, *cup9*Δ cells can import peptides in the absence of *UBR1*, whereas the import by *CUP9* cells requires *UBR1*. These findings (C. Byrd & A.V., unpublished data) strongly suggest that Cup9p is the postulated short-lived repressor which controls the rate of peptide import by regulating the expression of the Ptr2p transporter. Remarkably, an earlier study (Knight *et al.* 1994) identified Cup9p as a protein required for an aspect of resistance to copper toxicity in *S. cerevisiae*. Thus, one and the same physiological substrate of the N-end rule pathway functions as both a repressor of peptide import and a regulator of copper homeostasis.

### Gα subunit of G protein

Overexpression of the N-end rule pathway was found to inhibit the growth of haploid but not diploid cells (Madura & Varshavsky 1994). This ploidy-dependent toxicity was traced to the enhanced degradation of Gpa1p, the Gα subunit of the G protein that regulates cell differentiation in response to mating pheromone. The half-life of newly formed Gα at 30 °C is ≈50 min in wild-type cells, ≈10 min in cells overexpressing the N-end rule pathway, and >10 h in cells lacking the pathway. The degradation of Gα is preceded by its multiubiquitylation (Madura & Varshavsky 1994). Like other Gα subunits of G proteins, the *S. cerevisiae* Gpa1p bears a conjugated N-terminal myristoyl moiety, which appears to be retained on Gpa1p during its targeting for degradation. A deletion of the first 88 residues

of Gpa1p greatly accelerates its decay but retains the dependence of Gpa1p degradation on Ubr1p (K. Madura, unpublished data). These data suggest that Ubr1p recognizes a feature of Gα that is distinct from the N-degron. Another, N-degron-based model invokes a *trans*-targeting mechanism (Fig. 2C, D).

Physiological implications of the Ubr1p-dependent degradation of Gα remain to be understood. Because the metabolic stability of Gα is expected to be influenced by its functional state—Gα can be GTP- or GDP-bound, covalently modified, or associated with Gβγ, the pheromone receptor, and other Gα ligands, the degradation of Gα in yeast may function either to augment or to inhibit cell's responses to a pheromone. A $G_s$-type Gα is short-lived in mouse cells as well (Levis & Bourne 1992), consistent with the possibility that Gα subunits of other organisms are also degraded by the N-end rule pathway. The activation of mouse Gα shortens its *in vivo* half-life (Levis & Bourne 1992), suggesting an adaptation-related function of Gα degradation. Further, Obin *et al.* (1994) described the ATP-dependent degradation of all three subunits of the bovine retinal G protein in reticulocyte extract. (It is not known whether Gβ and/or Gγ subunits of the *S. cerevisiae* G protein are also metabolically unstable.) Hondermarck *et al.* (1992) (see also Taban *et al.* 1996) reported that differentiation of rat pheochromocytoma PC12 cells is inhibited by dipeptides bearing destabilizing N-terminal residues. (These compounds have been shown to inhibit the N-end rule pathway in *S. cerevisiae* (Baker & Varshavsky 1991); their efficacy as N-end rule inhibitors in mammalian cells remains to be evaluated.) Given the findings with Gα (Madura & Varshavsky 1994), one interpretation of these results (Hondermarck *et al.* 1992) is that inhibitors of the N-end rule pathway may suppress cell differentiation through a metabolic stabilization of the relevant Gα subunits in PC12 cells.

## Sindbis virus RNA Polymerase and other viral proteins

The Sindbis virus RNA polymerase, also called nsP4 (nonstructural Protein 4), is produced by an endoproteolytic cleavage of the viral precursor polyprotein nsP1234 (Strauss & Strauss 1994). The nsP4 protein bears N-terminal Tyr (a primary destabilizing residue; Figs 1 and 5A), and is degraded by the N-end rule pathway in reticulocyte extract (deGroot *et al.* 1991). Tyr is an N-terminal residue of other alphaviral RNA polymerases as well (Strauss & Strauss 1994), suggesting that these homologues of Sindbis polymerase are also degraded by the N-end rule pathway. Whereas the bulk of newly formed nsP4 is rapidly degraded, a fraction of nsP4 in infected cells is long-lived, presumably within a replication complex that contains viral and host proteins (Strauss & Strauss 1994, and references therein). This model may be generally applicable, in that physiological N-end rule substrates—including alphaviral RNA polymerases and Gα subunits of G proteins—are likely to exist in several states that differ by covalent modifications of a substrate and/or its associations with other ligands, and that consequently also differ by the rates at which various forms of a substrate are degraded by the N-end rule pathway.

There are many potential N-end rule substrates derived from viral polyproteins (Dougherty & Semler 1993). One of them is the integrase of the human immunodeficiency virus (HIV), produced by cleavages within the *gag-pol* precursor polyprotein. The processed integrase bears N-terminal Phe (Dougherty & Semler 1993), a strongly destabilizing residue in the N-end rule (Fig. 5A). Therefore it is possible that—similarly to the Sindbis virus RNA polymerase—at least a fraction of HIV integrase is short-lived *in vivo*.

## c-Mos, a proto-oncoprotein

This 39 kDa Ser/Thr-kinase is expressed predominantly in male and female germ cells. Sagata and colleagues have identified c-Mos as a physiological substrate of the N-end rule pathway that is targeted for degradation through its N-terminal Pro residue (Nishizawa *et al.* 1992, 1993). Met-Pro-Ser-Pro, the encoded N-terminal sequence of *Xenopus* c-Mos, is conserved among all vertebrates examined (Nishizawa *et al.* 1992). Since the N-terminal Met-Pro peptide bond is readily cleaved by the major cytosolic Met-aminopeptidases (Arfin & Bradshaw 1988), the initially second-position Pro is expected to appear at the N-terminus of nascent c-Mos cotranslationally or nearly so.

The activity of the Pro-based N-degron in c-Mos is inhibited through the phosphorylation of Ser-2 (Ser-3 in the c-Mos ORF) (Nishizawa *et al.* 1992, 1993). During the maturation of *Xenopus* oocytes, c-Mos is phosphorylated partially and reversibly, and therefore remains short-lived. Later, at the time of germinal vesicle breakdown and the arrest of mature oocytes (eggs) at the second meiotic metaphase, c-Mos becomes long-lived, owing to its nearly stoichiometric phosphorylation at Ser-2 (Watanabe *et al.* 1991). Fertilization or mechanical activation of a *Xenopus* egg releases the meiotic arrest through the induced degradation of c-Mos—caused by a nearly complete dephosphorylation of phosphoserine-2 (Nishizawa *et al.* 1992, 1993).

Consistent with this model of the N-degron in c-Mos, the replacement of Ser-2 with Asp or Glu (whose negative charge mimics that of the phosphoryl group) rendered c-Mos long-lived, whereas the replacement of Ser-2 with Ala yielded a constitutively unstable c-Mos (Nishizawa *et al.* 1992). Lys-33 (Lys-34 in the c-Mos ORF) is a major ubiquitylation site of the c-Mos N-degron (Nishizawa *et al.* 1993).

In contrast to N-terminal Pro in the context of c-Mos, the N-terminal Pro followed by the sequence His-Gly Ser-... (this is the context of engineered N-end rule substrates such as X-βgal and X-DHFR (Varshavsky 1992)) did not confer a short half-life on a reporter protein in either yeast or mammalian cells (F. Lévy, T. Rümenapf & A.V., unpublished data). One interpretation of these results is that the N-degron of c-Mos, whose conserved N-terminal sequence is Pro-Ser-Pro-..., has a 'degron-enabling' internal determinant additional to, and perhaps specific for, the N-terminal Pro. The c-Mos N-degron is the first example of N-degron whose activity is regulated by phosphorylation (Nishizawa *et al.* 1992).

### Compartmentalized proteins retrotransported to the cytosol

In contrast to cytosolic and nuclear proteins, the proteins that function in (or pass through) the ER, Golgi, and related compartments often bear destabilizing N-terminal residues—the consequence of cleavage specificity of signal peptidases, which remove signal sequences from proteins translocated into the ER (Bachmair *et al.* 1986). We have suggested that one function of the N-end rule pathway may be the degradation of previously compartmentalized proteins that 'leak' or are transported into the cytosol from compartments such as ER (Bachmair *et al.* 1986; Varshavsky 1992). Remarkably, it has recently been found that at least some compartmentalized proteins can be retrotransported to the cytosol through a route that requires specific ER proteins. US11, the ER-resident transmembrane protein encoded by cytomegalovirus, causes the newly translocated MHC class I heavy chain to be selectively retrotransported back to the cytosol, where the heavy chain is degraded by a proteasome-dependent pathway (Wiertz *et al.* 1996). Similarly, CPY*, a defective vacuolar carboxypeptidase of *S. cerevisiae*, is retrotransported to the cytosol shortly after entering the ER, and is degraded in the cytosol by a Ub/proteasome-dependent pathway that requires the Ubc7p Ub-conjugating enzyme (Hiller *et al.* 1996). The expected N-terminal residue of the translocated and processed MHC class I heavy chain is Gly—a stabilizing residue (Fig. 1). The expected N-terminal residue of the wild-type CPY carboxypeptidase whose signal sequence had been cleaved off is Ile—a primary destabilizing residue (Fig. 1). Whether the N-end rule pathway plays a role in the degradation of retrotransported proteins remains to be determined.

### Potential N-end rule substrates

Several cytosolic and nuclear proteins are known to bear destabilizing N-terminal residues, but have not been shown, thus far, to be degraded by the N-end rule pathway. Among them are the λ phage cII protein, the *S. cerevisiae* Cup1p protein, the catalytic subunits of calpains (calcium-dependent proteases), and several histone-like, micronucleus-specific proteins of *Tetrahymena*. These putative N-end rule substrates are discussed by Varshavsky *et al.* (1997).

### Applications of N-degron

The portability and modular organization of N-degrons make possible a variety of applications whose common feature is the conferring of a constitutive or conditional metabolic instability on a protein of interest. These applications are discussed elsewhere (Varshavsky 1995, 1996a,b).

## Concluding remarks

Although many things have been learnt about the N-end rule since its discovery 10 years ago, the answers to several key questions remain unknown. For example, the detailed mechanics of targeting is not understood. The biochemical dissection of the N-end rule pathway reconstituted *in vitro* from defined (cloned) components will be essential for attaining this goal. Crystallographic-quality structural information about N-recognin and the entire targeting complex will be required as well. The recently emerged possibility that N-recognin may target not only N-degrons but also other degradation signals adds yet another level of complexity which will have to be addressed.

Genetic screens for proteins degraded by the N-end rule pathway are our best hope for bringing to light physiological N-end rule substrates. It is already clear that at least some of these substrates are conditionally unstable—for example, partitioned between a short-lived free substrate and a long-lived complex of the substrate with other proteins. In addition, for some substrates, the rate-limiting step in their degradation

may be a processing (cleavage) event that produces an N-degron from a pre-N-degron. If so, a significant fraction of extant substrate molecules may bear a stabilizing N-terminal residue. Given these obstacles to identifying physiological N-end rule substrates, they are likely to be more numerous than is apparent at the present time.

## Acknowledgements

I am grateful to the current and former members of the laboratory, whose work on the N-end rule is described in this review. I thank colleagues whose names are cited in the text for their permission to discuss unpublished data. Our studies are supported by grants from the NIH and the Association for the Cure of the Cancer of the Prostate.

This article is a modified and updated version of an earlier review (Varshavsky 1996b).

## Abbreviations

Ub, ubiquitin; βgal, β-galactosidase; DHFR, dihydrofolate reductase; $Nt^N$-amidase, amidohydrolase specific for N-terminal Asn; R-transferase, Arg-tRNA-protein transferase; L/F-transferase, Leu, Phe-tRNA-protein transferase.

## References

Alagramam, K., Naider, F. & Becker, J.M. (1995) A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **15**, 225–234.

Arfin, S.M. & Bradshaw, S.A. (1988) Cotranslational processing and protein turnover in eukaryotic cells. *Biochemistry* **27**, 7979–7984.

Arnason, T.A. & Ellison, M.J. (1994) Stress resistance in *S. cerevisiae* is stronglycorrelated with assembly of a novel type of multiubiquitin chain. *Mol. Cell. Biol.* **14**, 7876–7883.

Baboshina, O.V. & Haas, A.L. (1996) Novel multiubiquitin chain linkages catalyzed by the conjugating enzymes E2EPF and RAD6 are recognized by 26S proteasome subunit 5. *J. Biol. Chem.* **271**, 2822–2831.

Bachmair, A., Finley, D. & Varshavsky, A. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186.

Bachmair, A. & Varshavsky, A. (1989) The degradation signal in a short-lived protein. *Cell* **56**, 1019–1032.

Baker, R.T. & Varshavsky, A. (1995) Yeast N-terminal amidase: a new enzyme and component of the N-end rule pathway. *J. Biol. Chem.* **270**, 12065–12074.

Baker, R.T. & Varshavsky, A. (1991) Inhibition of the N-end rule pathway in living cells. *Proc. Natl. Acad. Sci. USA* **88**, 1090–1094.

Balzi, E., Choder, M., Chen, W., Varshavsky, A. & Goffeau, A. (1990) Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**, 7464–7471.

Bartel, B., Wünning, I. & Varshavsky, A. (1990) The recognition component of the N-end rule pathway. *EMBO J.* **9**, 3179–3189.

Chau, V., Tobias, J.W., Bachmair, A. *et al.* (1989) A multiubiquitin chain is confined to a specific lysine in a targeted short-lived protein. *Science* **243**, 1576–1583.

Ciechanover, A., Ferber, S., Ganoth, D., Elias, S., Hershko, A. & Arfin, S. (1988) Purification and characterization of arginyl-tRNA-protein transferase from rabbit reticulocytes. *J. Biol. Chem.* **263**, 11155–11167.

Dayal, V.K., Chakraborty, G., Sturman, J.A. & Ingoglia, N.A. (1990) The site of amino acid addition to posttranslationally modified proteins of regenerating rat sciatic nerves. *Biochim. Biophys. Acta* **1038**, 172–177.

deGroot, R.J., Rümenapf, T., Kuhn, R.J., Strauss, E.G. & Strauss, J.H. (1991) Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc. Natl. Acad. Sci. USA* **88**, 8967–8971.

Dohmen, R.J., Wu, P. & Varshavsky, A. (1994) Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science* **263**, 1273–1276.

Dohmen, R.J., Madura, K., Bartel, B. & Varshavsky, A. (1991) The N-end rule is mediated by the Ubc2 (Rad6) ubiquitin-conjugating enzyme. *Proc. Natl. Acad. Sci. USA* **88**, 7351–7355.

Dougherty, W.G. & Semler, B.L. (1993) Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiol. Rev.* **57**, 781–822.

Finley, D., Bartel, B. & Varshavsky, A. (1989) The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. *Nature* **338**, 394–401.

Ghislain, M., Dohmen, R.J., Lévy, F. & Varshavsky, A. (1996) Cdc48p interacts with Ufd3p, a WD-repeat protein required for ubiquitin-dependent proteolysis in *Saccharomyces cerevisiae*. *EMBO J.* **15**, 4884–4899.

Gonda, D.K., Bachmair, A., Wünning, I., Tobias, J.W., Lane, W.S. & Varshavsky, A., (1989) Universality and structure of the N-end rule. *J. Biol. Chem.* **264**, 16700–16712.

Gottesman, S. & Maurizi, M.R. (1992) Regulation by proteolysis: energy-dependent proteases and their targets. *Microbiol. Rev.* **56**, 592–621.

Gottesman, S., Clark, W.P., de Crecy-Lagard, V. & Maurizi, M.R. (1993) ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*. *J. Biol. Chem.* **268**, 22618–22626.

Grigoryev, S., Stewart, A.E., Kwon, Y.T., Arfin, S.M., Bradshaw, R.A., Copeland, N.J. & Varshavsky, A. (1996) A mouse amidase specific for N-terminal asparagine: the gene, the enzyme, and their function in the N-end rule pathway. *J. Biol. Chem.* **271**, 28521–28532.

Hershko, A. (1991) The ubiquitin pathway for protein degradation. *Trends Biochem. Sci.* **16**, 265–268.

Hill, C.P., Johnston, N.L. & Cohen, R.E. (1993) Crystal structure of a ubiquitin- dependent degradation substrate: a three-disulfide form of lysozyme. *Proc. Natl. Acad. Sci. USA* **90**, 4136–4140.

Hiller, M.M., Finger, A., Scheiger, M. & Wolf, D.H. (1996) Endoplasmic reticulum associated degradation of a mutated soluble vacuolar enzyme, carboxypeptidase Y, occurs via the ubiquitin-proteasome pathway. *Science* **273**, 1725–1728.

Hondermarck, H., Sy, J., Bradshaw, R.A. & Arfin, S.M. (1992) Dipeptide inhibitors of ubiquitin-mediated protein turnover prevent growth factor-induced neurite outgrowth in rat

pheochromocytoma PC12 cells. *Biochem. Biophys. Res. Commun.* **189**, 280–288.

Jentsch, S. (1992) The ubiquitin-conjugation system. *Annu. Rev. Genet.* **26**, 179–207.

Johnson, E.S., Ma, P.C.M., Ota, I.M. & Varshavsky, A. (1995) A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.* **270**, 17442–17456.

Johnson, E.S., Gonda, D.K. & Varshavsky, A. (1990) *Cis-trans* recognition and subunit-specific degradation of short-lived proteins. *Nature* **346**, 287–291.

Johnston, J.A., Johnson, E.S., Waller, P.R.H. & Varshavsky, A. (1995) Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway, *J. Biol. Chem.* **270**, 8172–8178.

Kaiser, C.A., Preuss, D., Grisafi, P. & Botstein, D. (1987) Many random sequences functionally replace the secretion signal sequence of yeast invertase. *Science* **235**, 312–317.

Knight, S.A.B., Tamai, K.T., Kosman, D.J. & Thiele, D.J. (1994) Identification and analysis of a *Saccharomyces cerevisiae* copper homeostasis gene encoding a homeodomain protein. *Mol. Cell. Biol.* **14**, 7792–7804.

Knighton, D.R., Kan, C.C., Howland, E. *et al.* (1994) Structure of and kinetic channeling in bifunctional dihydrofolate reductase-thymidylate synthase, *Nature Struct. Biol.* **1**, 186–194.

Levis, M.J. & Bourne, H.R. (1992) Activation of the $\alpha$ subunit of $G_s$ in intact cells alters its abundance, rate of degradation, and membrane avidity. *J. Cell. Biol.* **119**, 1297–1307.

Lévy, F., Johnsson, N., Rümenapf, T. & Varshavsky, A. (1996) Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl. Acad. Sci. USA* **93**, 4907–4912.

Li, X. & Chang, Y.H. (1995) Amino-terminal protein processing in *Saccharomyces cerevisiae* is an essential function that requires two distinct methionine aminopeptidases. *Proc. Natl. Acad. Sci. USA* **92**, 12357–12361.

Madura, K. & Varshavsky, A. (1994) Degradation of G$\alpha$ by the N-end rule pathway, *Science* **265**, 1454–1458.

Madura, K., Dohmen, R.J. & Varshavsky, A. (1993) N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.* **268**, 12046–12054.

Maeda, T., Wurgler-Murphy, S.M. & Sato, H. (1994) A two-component system that regulates an osmosensing MAP kinase cascade in yeast. *Nature* **369**, 242–245.

Murray, A. & Hunt, T. (1993) *The Cell Cycle.* New York: W.H. Freeman & Co., pp. 60–62.

Negrutskii, B.S. & Deutscher, M.P. (1991) Channeling of aminoacyl-tRNA for protein synthesis *in vivo, Proc. Natl. Acad. Sci. USA* **88**, 4991–4995.

Nishizawa, M., Furuno, N., Okazaki, K., Tanaka, H., Ogawa, Y. & Sagata, N. (1993) Degradation of Mos by the N-terminal proline-dependent ubiquitin pathway on fertilization of *Xenopus* eggs: possible significance of natural selection for Pro-2 in Mos. *EMBO J.* **12**, 4021–4027.

Nishizawa, M., Okazaki, K., Furuno, N., Watanabe, N. & Sagata, N. (1992) The 'second-codon rule' and autophosphorylation govern the stability and activity of Mos during the meiotic cell cycle in *Xenopus* oocytes. *EMBO J.* **11**, 2433–2446.

Obin, M., Nowell, T. & Taylor, A. (1994) The photoreceptor G-protein transducin is a substrate for ubiquitin-dependent proteolysis. *Biochem. Biophys. Res. Communication* **200**, 1169–1176.

Ota, I.M. & Varshavsky, A. (1992) A gene encoding a putative tyrosine phosphatase suppresses lethality of an N-end rule-dependent mutant. *Proc. Natl. Acad. Sci. USA* **89**, 2355–2359.

Ota, I.M. & Varshavsky, A. (1993) A yeast protein similar to bacterial two-component regulators. *Science* **262**, 566–569.

Özkaynak, E., Finley, D., Solomon, M.J. & Varshavsky, A. (1987) The yeast ubiquitin genes: a family of natural gene fusions. *EMBO J.* **6**, 1429–1440.

Pahl, H.L. & Baeuerle, P.A. (1996) Control of gene expression by proteolysis. *Curr. Opin. Cell Biol.* **8**, 340–347.

Pickart, C. (1988) Ubiquitin carrier proteins. In: *Ubiquitin* (ed. M. Rechsteiner), pp. 77–99. Plenum Press, NY.

Rechsteiner, M., Hoffman, L. & Dubiel, W. (1993) The multicatalytic and 26S proteases. *J. Biol. Chem.* **268**, 6065–6068.

Reiss, Y., Kaim, D. & Hershko, A. (1988) Specificity of binding of N-terminal residue of proteins to ubiquitin-protein ligase. Use of amino acid derivatives to characterize specific binding sites. *J. Biol. Chem.* **263**, 2693–2698.

Sadis, S., Atienza, C. & Finley, D. (1995) Synthetic signals for ubiquitin-dependent proteolysis. *Mol. Cell. Biol.* **15**, 4086–4095.

Schatz, G. & Dobberstein, B. (1996) Common principles of protein translocation across membranes. *Science* **271**, 1519–1526.

Scheffner, M., Nuber, U. & Huibregtse, J.M. (1995) Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **373**, 81–83.

Sherman, F., Stewart, J.W. & Tsunasawa, S. (1985), Methionine or not methionine at the beginning of a protein? *BioEssays* **3**, 27–31.

Shrader, T.E., Tobias, J.W. & Varshavsky, A. (1993) The N-end rule in *Escherichia coli*: cloning and analysis of the leucyl, phenylalanyl-tRNA-protein transferase gene *aat*. *J. Bact.* **175**, 4364–4374.

Simon, S.M. & Blobel, G. (1991) A protein-conducting channel in the endoplasmic reticulum. *Cell* **69**, 371–380.

Spence, J., Sadis, S., Haas, A.L. & Finley, D. (1995) A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol. Cell Biol.* **15**, 1265–1273.

Stewart, A.E., Arfin, S.M. & Bradshaw, R.A. (1995) The sequence of porcine protein N-terminal asparagine amidohydrolase: a new component of the N-end rule pathway. *J. Biol. Chem.* **270**, 25–28.

Strauss, J.H. & Strauss, E.G. (1994) The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* **58**, 491–562.

Taban, C.H., Hondermarck, H., Bradshaw, R.A. & Biolly, B. (1996) Effect of a dipeptide inhibiting ubiquitin-mediated protein degradation nerve-dependent limb regeneration in the newt. *Experientia* **52**, 865–70.

Tobias, J.W., Shrader, T.E., Rocap, G. & Varshavsky, A. (1991) The N-end rule in bacteria. *Science* **254**, 1374–1377.

Varshavsky, A., Byrd, C., Davydov, I.V. *et al.* (1997) The N-end rule pathway. In: *Ubiquitin and the Biology of the Cell*, (eds D. Finley & J.-M. Peters) Plenum Press, NY. (in press).

Varshavsky, A. (1996a) The N-end rule. *Cold Spring Harbor Symp. Quant. Biol.* **60**, 461–478.

Varshavsky, A. (1996b) The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* **93**, 12142–12149.

Varshavsky, A. (1995) Codominance and toxins: a path to drugs

of nearly unlimited selectivity. *Proc. Natl. Acad. Sci. USA* **92**, 3663–3667.

Varshavsky, A. (1992) The N-end rule. *Cell* **69**, 725–735.

Varshavsky, A. (1991) Naming a targeting signal. *Cell* **64**, 13–15.

Watanabe, N., Hunt, T., Ikawa, Y. & Sagata, N. (1991) Independent inactivation of MPF and cytostatic factor (Mos) upon fertilization of *Xenopus* eggs. *Nature* **352**, 247–248.

Watkins, J.F., Sung, P., Prakash, L. & Prakash, S. (1993) The *Saccharomyces cerevisiae* DNA repair gene *RAD23* encodes a nuclear protein containing ubiquitin-like domain required for biological function. *Mol. Cell. Biol.* **13**, 7757–7765.

Wawrzynov, A., Wojtkowiak, D., Marzalec, J., *et al.* (1995) The ClpX heat-shock protein of *Escherichia coli*, the ATP-dependent specificity component of the ClpP-ClpX protease, is a novel molecular chaperone. *EMBO J.* **14**, 1867–1877.

Wickner, S., Gottesman, S., Skowyra, D., Hoskins, J., McKenney, K. & Maurizi, M.R. (1994) A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc. Natl. Acad. Sci.* USA **91**, 12218–12222.

Wiertz, E.J.H.J., Jones, T.R., Sun, L., Bogyo, M., Geuze, H.J. & Ploegh, H.L. (1996) The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol. *Cell* **84**, 769–779.

Wolf, S., Lottspeich, F. & Baumeister, W. (1993) Ubiquitin found in the archaebacterium *Thermoplasma acidophilum*. *FEBS Lett.* **326**, 42–44.

# The ubiquitin system

Alexander Varshavsky

Eukaryotes contain a highly conserved multi-enzyme system that covalently links ubiquitin to a variety of intracellular proteins that bear degradation signals recognized by this system. The resulting ubiquitin–protein conjugates are degraded by the 26S proteasome, a large ATP-dependent protease. Pathways that involve ubiquitin underlie a multitude of processes, including cell differentiation, the cell cycle and responses to stress.

**UBIQUITIN (Ub)** is a 76-residue protein that exists in cells either free or covalently linked to other proteins. Ub-dependent pathways have been shown to play major roles in a legion of biological processes, including cell differentiation, the cell cycle, embryogenesis, apoptosis, signal transduction, DNA repair, transmembrane and vesicular transport, stress responses (including the immune response) and functions of the nervous system. Many, but not all, of the Ub-dependent pathways involve processive degradation of Ub-conjugated (ubiquitylated*) proteins by the 26S proteasome – an ATP-dependent, multisubunit protease. I will remark on the history of this field, consider the enzymology and mechanics of the Ub system, and describe some of its degradation signals and physiological functions. A review of this length can offer but a glimpse of the Ub system, making partial amends with references to more detailed reviews and recent original works.

## The beginnings

Ub was first described in 1975 as an abundant, highly conserved protein, hence its name[1]. In 1978–1985, Hershko and co-workers identified Ub as an essential component of the ATP-dependent proteolytic system in an extract from rabbit reticulocytes, and deciphered *in vitro* the E1–E2–E3 enzymology of Ub conjugation[2]. Having been attained in the setting of a cell-free system, these important early advances did not reveal the functions of the ubiquitin system. They also did not address the nature of a degradation signal in a short-lived protein.

**A. Varshavsky** is in the Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.
Email: avarsh@cco.caltech.edu

In 1984, we identified the mouse cell line ts85 as a temperature-sensitive mutant in the Ub-activating enzyme[3], and produced the first evidence that Ub is required for protein degradation in living cells. Ub-dependent proteolysis was found to be essential for cell-cycle progression, as ts85 cells were arrested specifically in the G2 phase of the cell cycle[3]. This insight presaged the understanding of the role of cyclins and cyclin-like proteins whose periodic Ub-dependent destruction drives and regulates the cell cycle[4,5].

In 1986, we discovered the first degradation signal of the Ub system. The *in vivo* half-life of a protein was found to depend on the identity of the protein's N-terminal residue – a relation termed the N-end rule[6,7].

In 1987, following the initial insight with ts85 cells[3], we discovered the first physiological functions of the Ub system, by demonstrating that RAD6 and CDC34 – key components of DNA repair and cell cycle control – are in fact Ub-conjugating enzymes[8,9]. A non-proteolytic, chaperonin function of Ub was also identified – by the finding that a transient linkage of Ub to nascent ribosomal proteins is required for the efficient biogenesis of ribosomes[10].

In 1989, we discovered the multi-Ub chain[11], an advance that initiated the understanding of the mechanistic role of Ub. It was also found that Ub-dependent proteolysis can destroy a subunit of an oligomeric protein *selectively*, leaving intact the rest of the protein's subunits[12]. This fundamental capability of the Ub system accounts for large differences in the *in vivo* half-lives of subunits in many regulatory proteins.

Another advance, which concluded the early era of Ub studies, was the demonstration by several laboratories that a ~2000 kDa, ATP-dependent protease, termed the 26S proteasome, specifically degrades Ub–protein conjugates[2,13–15]. The 26S proteasome consists of the previously identified 20S 'core' proteasome[16–18] and a complex containing multiple ATPases (the 19S 'cap') at both ends of the 20S proteasome[19–21].

## Organization and enzymology of the ubiquitin system

Eukaryotic cells contain Ub-specific enzymes that catalyse reactions whose product is either a single Ub moiety or a multi-Ub chain covalently linked to an acceptor protein (Fig. 1). Ub is conjugated to other proteins through an amide bond, called the isopeptide bond, between the C-terminal (Gly76) residue of Ub and the ε-amino group of a Lys residue in an acceptor protein[2,7,21–23].

Ub is activated for conjugation to other proteins by a Ub-activating enzyme (E1), which couples ATP hydrolysis to the formation of a high-energy thioester bond between Gly76 of Ub and a specific Cys residue of E1 (Fig. 1)[2,21,23]. The E1-linked Ub moiety is moved, in a transesterification reaction, from E1 to a Cys residue of a Ub-conjugating enzyme (E2), and from there to a Lys residue of an ultimate acceptor protein, yielding a Ub–protein conjugate (Fig. 1). This last step requires the participation of another component, called E3 or recognin[7,21,23], which selects a protein for ubiquitylation through an interaction with its degradation signal. At least three of the known E3 proteins – the mammalian E6AP, its yeast (*Schizosaccharomyces pombe*) homolog Pub1p, and Ubr1p, the E3 of the N-end rule pathway – are also bona fide enzymes that act at the step between an E2 and an ultimate acceptor of Ub. These (and possibly other) E3s catalyse in particular the movement, through transesterification, of the Ub moiety from the Cys residue of a relevant (E3-associated) E2 enzyme to a Cys residue of E3 itself – in a reaction analogous to the Ub transfer from E1 to E2 (Fig. 1) (see Refs 24, 25; V. Chau and A. Varshavsky, unpublished).

**Multi-ubiquitin chains.** In these structures, the C-terminal Gly of one Ub is joined to an internal Lys of the adjacent Ub moiety, resulting in a chain of Ub–Ub conjugates containing two or more Ub moieties[11,26]. Ub has seven lysines; thus,

*Ubiquitin whose C-terminal (Gly76) carboxyl group is covalently linked to another compound is called the *ubiquityl* moiety, the derivative terms being *ubiquitylation* and *ubiquitylated*. The term Ub refers to both free ubiquitin and the ubiquityl moiety. This nomenclature[67], which is also recommended by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology[68], brings ubiquitin-related terms in line with the standard chemical terminology.

 0968–0004/97/$17.00 PII: S0968-0004(97)01122-5

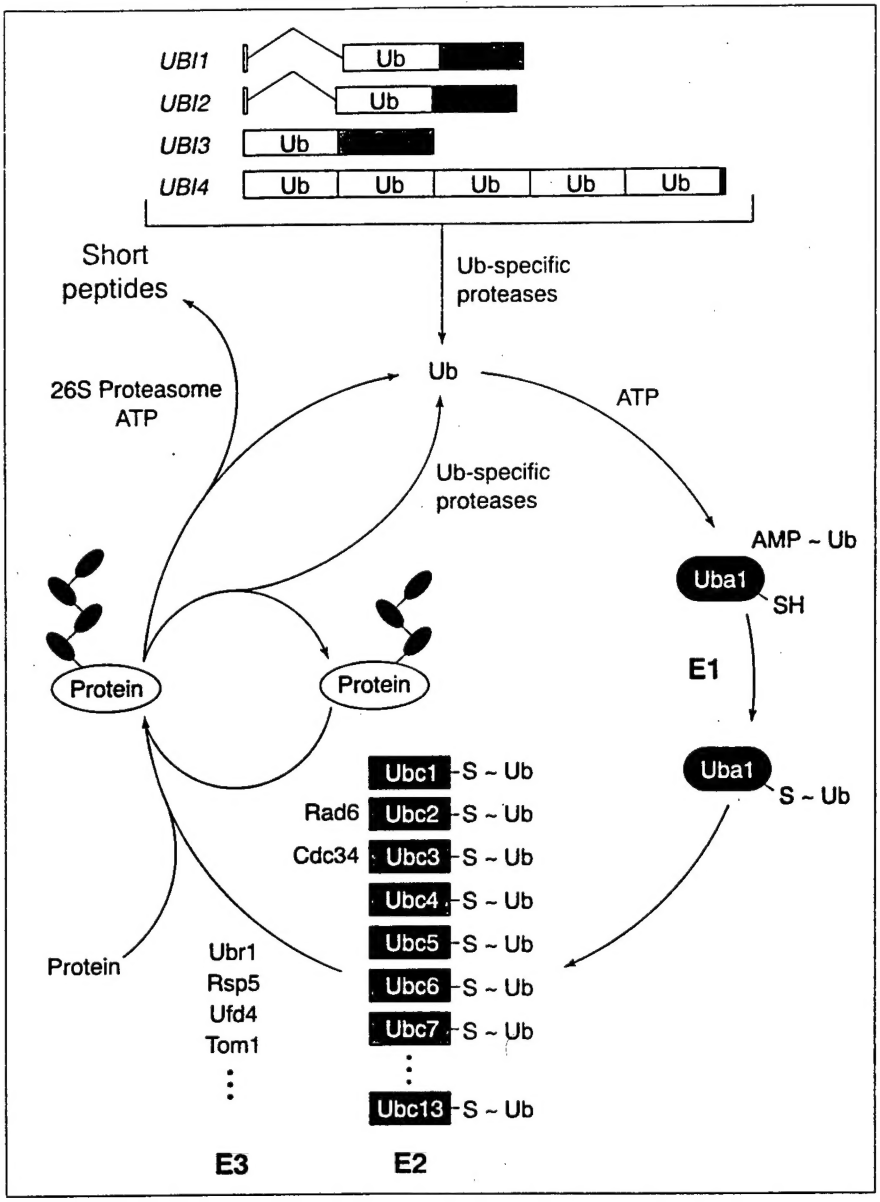



**Figure 1**

The ubiquitin system of *Saccharomyces cerevisiae*. The yeast ubiquitin (Ub) genes, two of which (*UBI1* and *UBI2*) contain introns, encode fusions of Ub (yellow rectangles) to itself (*UBI4*) or to one of the two specific ribosomal proteins (*UBI1–UBI3*) (red and blue rectangles)[12,69]. These fusions are cleaved by Ub-specific processing proteases[21], yielding mature Ub. Thioester bonds between Ub and the active-site Cys residues of Ub-specific enzymes are denoted by '~'. The conjugation of Ub to other proteins involves a preliminary ATP-dependent step, in which the last (Gly76) residue of Ub is joined, via a thioester bond, to a Cys residue in the Ub-activating (E1) enzyme encoded by the *UBA1* gene[70]. The activated Ub is transferred to a Cys residue in one of at least 13 distinct Ub-conjugating (E2) enzymes encoded by the genes *UBC1–UBC13* (Refs 8, 21, 23), and from there to a Lys residue of an ultimate acceptor protein (yellow oval). This last step, and apparently also the formation of a multi-Ub chain (black ovals), requires the participation of another component, called E3 or recognin[7,21]. The function of E3 includes, but is not limited to, the recognition of a degradation signal in the acceptor protein (see the main text). The names of some of the currently known yeast E3s (Ubr1p, Rsp5p, Ufd4p and Tom1p)[21] are indicated as well. 'Rad6' and 'Cdc34' refer to the alternative (earlier) names of the Ubc2p and Ubc3p E2 enzymes[23]. A targeted, ubiquitylated protein is processively degraded to short peptides by the ATP-dependent 26S proteasome (see the main text).

a multi-Ub chain in which no Ub moiety has more than one Lys residue linked to another Ub moiety can have, a priori, any of the seven 'pure' topologies or a far greater number of 'mixed' topologies. The first multi-Ub chain to be discovered had its Ub moieties conjugated through the Lys48 residue of Ub[7,11]. Other multi-Ub chains involve Lys63 or Lys29 of Ub[27–29]. A chain linked through the Lys63 residues of its Ub moieties appears to have a distinct role in a pathway of DNA repair[27]. As discussed below, one function of a substrate-linked multi-Ub chain is to facilitate the substrate's degradation by the 26S proteasome.

**De-ubiquitylation.** The covalent bond between Ub and other proteins can be cleaved: there are multiple, ATP-independent proteases (*Saccharomyces cerevisiae* has at least 20 of them) whose common property is the ability to recognize a Ub moiety and cleave at the Ub-adduct junction[21,30,31]. One cause of the striking multiplicity of Ub-specific processing proteases (UBPs) is the diversity of their targets, which include linear (DNA-encoded) Ub fusions[10], Ub adducts with small nucleophiles such as glutathione[32], and also free[27] or substrate-linked multi-Ub chains[11,27–29].

The junctions in linear Ub adducts (natural or engineered Ub fusions) are structurally distinct from the junctions in branched Ub conjugates. Thus, some of the UBPs may have preferences for either linear or branched Ub adducts. In addition, a Ub conjugate can be spatially confined in a cell, making it accessible only to some UBPs. For example, a UBP associated with the 26S proteasome may have privileged access to the multi-Ub chain of a proteasome-bound substrate. Recent evidence suggests that this is indeed the case[2,21], and that proteasome-associated UBPs can function as editing enzymes, whose ability to de-ubiquitylate an already targeted (proteasome-bound) Ub–substrate conjugate can modulate the rate of degradation of the substrate[33].

A UBP may also bind specifically to a protein that bears a degradation signal. In the absence of this UBP, the protein is ubiquitylated and degraded, whereas the presence of the protein-bound UBP prevents the protein's degradation – by counteracting the formation of a protein-linked multi-Ub chain. Potential examples of this UBP function in physiologically relevant settings have recently been discovered[34–36], bringing to light yet another way in which the Ub system regulates the destruction of specific proteins.

## Degradation signals of the Ub system

Degradation signals are features of proteins that confer metabolic instability. The number of distinct Ub-dependent degradation signals (defined as the number of different, functionally non-redundant targeting complexes) in, for example, *S. cerevisiae*, is unknown, but is likely to exceed ten. Some of these signals are understood much better than

the others, and new signals are certain to be discovered. Degradation signals can be active constitutively or conditionally. Signals of the latter class – found in many regulators, including cyclins and transcription factors – are controlled through phosphorylation or interactions with other proteins, whose binding may sterically shield an otherwise constitutive degradation signal[5,7,21,37]. Alternatively, a protein, by forming a complex with another protein, may confer a short *in vivo* half-life on the latter – through the recruitment of a specific targeting complex of the Ub system. One physiologically relevant example of this mechanism is the enhancement of instability of the short-lived p53 transcriptional regulator (and tumor suppressor) by the papilloma virus E6 protein, whose binding to p53 recruits a targeting complex containing in particular the E6AP E3 protein[24,38].

**The N-degron.** This degradation signal was discovered through the invention of the Ub fusion technique, which made it possible to produce any desired residue at the N-terminus of a test protein in a cell[6]. The *in vivo* half-life of a protein was found to depend on the identity of its N-terminal residue – a relation termed the N-end rule[6,7]. In eukaryotes, the N-end rule pathway is a part of the Ub system. However, a version of this pathway is also present in bacteria such as *Escherichia coli*, which lack the Ub system[7]. The N-degron, a signal recognized by the N-end rule pathway, comprises two essential determinants: a destabilizing N-terminal residue and one or more internal lysines of a substrate[12,39]. The Lys residue is the site of formation of a multi-Ub chain[11]. The N-end rule pathway is reviewed in Ref. 7.

Subsequently identified degradation signals appear to be organized similarly to the N-degron, in that they comprise two distinct elements – an amino acid sequence or a conformational determinant (analogous to a destabilizing N-terminal residue) and a Lys residue or residues, the latter being the site of ubiquitylation.

**N-terminal ubiquitin as a degradation signal.** Protein-linked multi-Ub chains function as secondary signals for degradation, in that post-translational conjugation of Ub to a substrate is mediated by the substrate's amino acid sequences that act as a primary degradation signal. An exception to this arrangement is a linear (DNA-encoded) Ub fusion bearing a nonremovable N-terminal Ub moiety; a protein thus designed is short-lived in *S. cerevisiae*, its Ub functioning as a primary degradation signal[29]. The normally rapid cleavage of linear Ub fusions at the Ub–protein junction can be abolished (making the Ub moiety 'nonremovable') by converting the last (Gly) residue of Ub into Val[29]. Such Ub fusions are degraded by the Ub/proteasome-dependent system termed the UFD pathway (ubiquitin/fusion/degradation)[29], which is distinct from the N-end rule pathway[29,40].

What are the physiological UFD substrates? Eukaryotes contain a number of genes encoding proteins that bear domains highly similar to Ub[21]. Recently, one such protein, Rad23p of *S. cerevisiae*[41], was found to be conditionally short-lived, its degradation being mediated by the UFD pathway and the Ub-like N-terminal domain of Rad23p (K. Madura, pers. commun.). Because this domain is required for the Rad23p function of repairing UV-damaged DNA[41], it is likely that the metabolic instability of Rad23p is a significant aspect of its function. This first example of a physiological UFD substrate suggests that Ub-like N-terminal domains of other proteins may also act as degradation signals, and that metabolic instability is relevant to the functions of proteins bearing Ub-like domains.

**Other degradation signals.** One of the better-understood signals resides in the transcriptional activator Gcn4p, which induces *S. cerevisiae* genes whose products mediate the synthesis of amino acids and purines. Gcn4p is a short-lived protein ($t_{1/2} \approx 5$ min) whose degradation requires the Cdc34p Ub-conjugating (E2) enzyme[42]. The sequence SSSTDSTP (in single-letter abbreviations) at positions 99–106 of the 281-residue Gcn4p encompasses a major part of its degradation signal. (Conversions of certain residues in this sequence into alanines significantly stabilized Gcn4p; a deletion of the entire eight-residue region had the strongest effect, increasing the $t_{1/2}$ of Gcn4p by approximately tenfold[42].)

Sequences rich in Pro, Glu, Ser and Thr, referred to as PEST, have been proposed to function as degradation signals[43]. PEST-like sequences were indeed implicated in the Cdc34p-dependent degradation of Gcn4p and a cyclin Cln3p (Refs 42, 44). At the same time, the PEST motifs as currently formulated are present in roughly one-third of the known open reading frames[5]. To serve as useful prognosticators of degradation signals, these motifs will have to be defined more selectively.

No single Lys residue in the vicinity of the 99–106 sequence was uniquely required for the Gcn4p degradation, but converting five or more of these lysines into arginines (which cannot be ubiquitylated[21]) did cause a significant stabilization of Gcn4p (Ref. 42). This result is consistent with the previously proposed 'stochastic-capture' mechanism, in which a reversible binding of E3 to a non-Lys determinant of a degradation signal initiates a time-restricted search by the E3–E2 complex for a sterically accessible Lys residue[7,39]. An E3–E2 complex that recognizes the degron of Gcn4p contains in particular Cdc53p, which is bound to the Cdc34p E2 enzyme and is required for the degradation of *S. cerevisiae* G1 cyclins and cyclin-like proteins, such as Cln2p (Ref. 45) and Sic1p. Homologs of Cdc53p were identified in all eukaryotes examined, suggesting that a complex containing Cdc53p, Cdc34p and other proteins[45,46] targets phylogenetically conserved sequence motifs present in many short-lived proteins and exemplified by the degradation signal of Gcn4p.

Matα2, the *S. cerevisiae* cell-type specific transcriptional repressor, has two distinct degradation signals[47], at least one of which has the unusual property of requiring two E2 enzymes, Ubc6p and Ubc7p, for its activity[21]. A mammalian transcription factor c-Jun is a short-lived protein whose Ub-dependent degradation signal is located in a region called δ-domain[21,37]. The degradation of c-Jun is inhibited upon its functional activation by the mitogen-activated protein kinase (MAPK), which phosphorylates Ser and Thr residues in the vicinity of the δ-domain[48]. NF-κB, another mammalian transcription factor, is regulated in part through an association with IκBα, a conditionally short-lived subunit of the functionally inactive, cytosolic complex containing NF-κB[37,49,50]. Phosphorylation of IκBα, in response to a variety of stresses, activates its degradation signal, resulting in the Ub-dependent, subunit-selective degradation of IκBα that releases the previously cytosolic NF-κB for translocation into the nucleus. Another form of inactive (cytosolic) NF-κB contains an inhibitory 105-kDa subunit whose Ub-dependent *partial* degradation yields the p50 subunit and results in the active (nuclear) form of NF-κB. Thus, the Ub system is capable of destroying, processively, a specific *region* of a polypeptide while sparing the rest of it[37]. (This simplified description of the NF-κB system belies its great complexity[49,50].)

**Folding vs degradation.** Most (but not all) of the damaged or otherwise abnormal proteins are recognized and destroyed by the Ub system – apparently through the

exposure of their normally buried degradation signals. This raises the possibility that a normal and otherwise long-lived but nascent (and therefore conformationally immature) protein may be 'regarded' by the Ub system as damaged until the protein assumes its final conformation. Thus, the folding of at least some newly formed proteins is expected to be in kinetic competition with pathways that target them for degradation. Chaperonins, including members of the Hsp70p family, play major roles not only in the folding of newly made proteins but also in the competing process of their degradation by the Ub system. Understanding of the interplay between the folding and proteolytic pathways has markedly advanced in recent years[51–53].

### Mechanism of ubiquitin-dependent proteolysis

Why must a multi-Ub chain be linked to a short-lived protein before its destruction by the 26S proteasome? We proposed that a major function of a substrate-linked multi-Ub chain is to decrease the rate of dissociation of a substrate–proteasome complex, thereby increasing the probability of substrate degradation[11,54]. Suppose that a rate-limiting step, which leads to the first proteolytic cleavage of the proteasome-bound substrate, is the unfolding of a relevant region of the substrate – driven by thermal fluctuations and/or ATPases of the 26S proteasome. If so, an increase in stability of the proteasome–substrate complex, brought about by the multi-Ub chain, should facilitate substrate degradation, because the longer the allowed 'waiting' time, the greater the probability of a required unfolding event.

Identification of proteasome subunits that bind multi-Ub chains[55,56] is consistent with this model. Furthermore, a 17-kDa, lysine-lacking (and therefore non-ubiquitylatable) fragment of the 70-kDa Sindbis virus polymerase is still degraded by the N-end rule pathway (T. Rümenapf, J. Strauss and A. Varshavsky, unpublished). Thus, the ubiquitylation requirement characteristic of the previously studied N-end rule substrates may be a consequence of their relatively stable conformations. In other words, the delivery of a largely unfolded substrate to active sites in the interior of the proteasome may be so fast that the dissociation-slowing function of the multi-Ub chain is no longer required.

A physiologically relevant example of ATP-dependent, Ub-*independent* proteolysis by the 26S proteasome is the degradation of ornithine decarboxylase (ODC) complexed with antizyme, an ODC-binding protein induced by polyamines[57]. The Ub-independent proteolysis of ODC requires the presence of the ODC-bound antizyme, which remains intact in the course of ODC degradation[57]. The binding of antizyme to ODC may produce an active degradation signal in ODC. In addition, the antizyme moiety of the ODC–antizyme complex may specifically bind to the 26S proteasome, in which case it could play a role analogous to that of a substrate-linked multi-Ub chain.

### Physiological functions of the ubiquitin system

The effect of a given intracellular protein on the rest of the cell is determined in part by the concentration of active protein molecules. This parameter, in turn, is determined by the rate of synthesis of the protein in relation to the rate of its degradation, inactivation by other means or export from the compartment. The vast functional range of the Ub system stems from the diversity of its physiological substrates, whose number in a cell is likely to be comparable to the number of substrates of phosphokinases. In other words, it is the constitutive or conditional degradation of many specific proteins (cyclins, transcription factors, components of signal transduction pathways, damaged proteins) by the Ub system that underlies the diversity of its roles, some of which are described or alluded to above, as is also the non-proteolytic, chaperonin function of Ub[10]. Considered below are two other functions, among the many that are already known[7,21,37,38,50,58].

**Ubiquitin as a signal for endocytosis.** Proteins embedded in the plasma membrane are removed into the cell interior through endocytosis. A fraction of the removed protein is eventually returned to the plasma membrane, whereas the rest is delivered to the lysosome/vacuole and degraded there by vacuolar proteases. Hicke and Riezman[59] have shown that the endocytosis of *S. cerevisiae* Ste2p, an integral membrane protein and receptor for the α-factor pheromone, requires ubiquitylation of a specific Lys residue in the cytosol-exposed C-terminal region of Ste2p. This lysine is located in the sequence SINNDAKSS, which is sufficient for receptor internalization. Other examples of this new Ub function have also been described[60–62]. It remains to be understood why, in these cases, a multi-Ub chain linked to a cytosol-exposed region of a transmembrane protein signals endocytosis rather than degradation by the 26S proteasome.

**Degradation of compartmentalized proteins retrotransported to the cytosol.** It has been shown that proteins translocated from the cytosol into the endoplasmic reticulum (ER) may, under certain conditions, be retrotransported back to the cytosol, apparently through the same (bidirectional) channel in the ER membrane[63–66]. At least some of the retrotransported proteins are degraded in the cytosol by the Ub system, whose physiological substrates thus include compartmentalized proteins as well. A major class of targets for retrotransport are proteins that fail to fold or oligomerize properly in the ER. For example, a defective vacuolar carboxypeptidase of *S. cerevisiae* is retrotransported to the cytosol shortly after entering the ER, and is degraded by a Ub–proteasome-dependent pathway that requires the Ubc7p E2 enzyme[63]. Viruses have been shown to take advantage of the retrotransport machinery to manipulate the immune response of the infected host. Two proteins of the human cytomegalovirus, US11 and US2, cause the newly translocated class I heavy chain of the major histocompatibility complex to be selectively retrotransported back to the cytosol, where the heavy chain is destroyed by a proteasome-dependent pathway[64].

Many compartmentalized proteins, which function in (or pass through) the ER, Golgi and related compartments bear N-terminal residues that are recognized as destabilizing by the Ub-dependent N-end rule pathway[6,54]. In contrast, most of the cytosolic and nuclear proteins bear either stabilizing N-terminal residues or blocked (acetylated) amino termini. This difference is caused in part by the cleavage specificity of signal peptidases, which remove signal sequences from proteins translocated into the ER. We suggested that one function of the N-end rule pathway may be the degradation of previously compartmentalized proteins that return to the cytosol, by any route, from compartments such as the ER[6]. Whether the N-end rule pathway mediates the degradation of at least some retrotransported proteins remains to be determined.

### Concluding remarks

From its beginnings in the 1980s, the field of Ub studies has grown enormously and became relevant to many areas of biology. The deepening understanding of the Ub system is likely to result, among other things, in powerful experimental tools for manipulating the *in vivo* half-lives of intracellular proteins whose malfunction or overproduction leads to

cancer and other diseases. This approach to pharmacological intervention should further enlarge the already immense scope of Ub studies.

**Acknowledgements**

I thank D. Finley, S. Jentsch and K. Madura for helpful discussions, and G. Turner, L. Peck and A. Webster for comments on the manuscript.

**References**

1 Schlesinger, D. H., Goldstein, G. and Niall, H. D. (1975) *Biochemistry* 14, 2214–2218
2 Hershko, A. (1996) *Trends Biochem. Sci.* 21, 445–449
3 Finley, D., Ciechanover, A. and Varshavsky, A. (1984) *Cell* 37, 43–55
4 Murray, A. and Hunt, T. (1993) *The Cell Cycle*, Freeman
5 King, R. W., Deshaies, R. J., Peters, J. M. and Kirschner, M. W. (1996) *Science* 274, 1652–1659
6 Bachmair, A., Finley D. and Varshavsky, A. (1986) *Science* 234, 179–186
7 Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 12142–12149
8 Jentsch, S., McGrath, J. P. and Varshavsky, A. (1987) *Nature* 329, 131–134
9 Goebl, M. G. *et al.* (1988) *Science* 241, 1331–1335
10 Finley, D., Bartel, B. and Varshavsky, A. (1989) *Nature* 338, 394–401
11 Chau, V. *et al.* (1989) *Science* 243, 1576–1583
12 Johnson, E. S., Gonda, D. K. and Varshavsky, A. (1990) *Nature* 346, 287–291
13 Rechsteiner, M., Hoffman, L. and Dubiel, W. (1993) *J. Biol. Chem.* 268, 6065–6068
14 Hilt, W. and Wolf, D. H. (1996) *Trends Biochem. Sci.* 21, 96–102
15 Coux, O., Tanaka, K. and Goldberg, A. L. (1996) *Annu. Rev. Biochem.* 65, 801–817
16 Orlowski, M. and Wilk, S. (1981) *Biochem. Biophys. Res. Commun.* 101, 814–822
17 Groll, M. *et al.* (1997) *Nature* 386, 463–471
18 Lupas, A., Flanagan, J. M., Tamura, T. and Baumeister, W. (1997) *Trends Biochem. Sci.* 22, 399–404
19 Gray, C. W., Slaughter, C. A. and DeMartino, G. N. (1994) *J. Mol. Biol.* 236, 7–15
20 Jentsch, S. and Schlenker, S. (1995) *Cell* 82, 881–884
21 Hochstrasser, M. (1996) *Annu. Rev. Genet.* 30, 405–439
22 Ciechanover, A. S. (1994) *Cell* 79, 13–21
23 Jentsch, S. (1992) *Annu. Rev. Genet.* 26, 179–207
24 Scheffner, M., Nuber, U. and Huibregtse, J. M. (1995) *Nature* 373, 81–83
25 Nefsky, B. and Beach, D. (1996) *EMBO J.* 15, 1301–1312
26 Beal, R. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 861–866
27 Spence, J., Sadis, S., Haas, A. L. and Finley, D. (1995) *Mol. Cell. Biol.* 15, 1265–1273
28 Arnason, T. A. and Ellison, M. J. (1994) *Mol. Cell. Biol.* 14, 7876–7883
29 Johnson, E. S., Ma, P. C. M., Ota, I. M. and Varshavsky, A. S. (1995) *J. Biol. Chem.* 270, 17442–17456
30 Wilkinson, K. D. *et al.* (1995) *Biochemistry* 34, 14535–14546
31 Baker, R. T., Tobias, J. W. and Varshavsky, A. (1992) *J. Biol. Chem.* 267, 23364–23375
32 Pickart, C. M. and Rose I. A. (1985) *J. Biol. Chem.* 260, 7903–7910
33 Lam, Y. A., Xu, W., DeMartino, G. N. and Cohen, R. E. (1997) *Nature* 385, 737–740
34 Huang, Y., Baker, R. T. and Fischer-Vize, J. A. (1995) *Science* 270, 1828–1831
35 Moazed, D. and Johnson, A. D. (1996) *Cell* 86, 667–677
36 Everett, R. D. *et al.* (1997) *EMBO J.* 16, 566–577
37 Pahl, H. L. and Baeuerle, P. A. (1996) *Curr. Opin. Cell Biol.* 8, 340–347
38 Scheffner, M. *et al.* (1990) *Cell* 63, 1129–1136
39 Bachmair, A. and Varshavsky, A. (1989) *Cell* 56, 1019–1032
40 Ghislain, M., Dohmen, R. J., Levy, F. and Varshavsky, A. (1996) *EMBO J.* 15, 4884–4899
41 Watkins, J. F., Sung, P., Prakash, L. and Prakash, S. (1993) *Mol. Cell. Biol.* 13, 7757–7765
42 Korn, D., Raboy, B., Kulka, R. G. and Fink, G. R. (1994) *EMBO J.* 13, 6021–6030
43 Rechsteiner, M. and Rogers, S. W. (1996) *Trends Biochem. Sci.* 21, 267–271
44 Yaglom, J. *et al.* (1995) *Mol. Cell. Biol.* 15, 731–741
45 Willems, A. R. *et al.* (1996) *Cell* 86, 453–463
46 Bai, C. *et al.* (1996) *Cell* 86, 263–274
47 Hochstrasser, M. and Varshavsky, A. (1990) *Cell* 61, 697–708
48 Musti, A. M., Treier, M. and Bohmann, D. (1997) *Science* 275, 400–402
49 Baeuerle, P. A. and Baltimore, D. (1996) *Cell* 87, 13–20
50 Thanos, D. and Maniatis, T. (1995) *Cell* 80, 529–532
51 Gottesman, S., Wichkner, S. and Maurizi, M. R. (1997) *Genes Dev.* 11, 815–823
52 Hartl, F. U. (1996) *Nature* 381, 571–580
53 Suzuki, C. K. *et al.* (1997) *Trends Biochem. Sci.* 22, 118–123
54 Varshavsky, A. (1992) *Cell* 69, 725–735
55 Deveraux, Q., Ustrell, V., Pickart, C. and Rechsteiner, M. (1994) *J. Biol. Chem.* 269, 7059–7061
56 van Nocker, S. *et al.* (1996) *Mol. Cell. Biol.* 16, 6020–6028
57 Tokunaga, F. *et al.* (1994) *J. Biol. Chem.* 269, 17382–17385
58 Hedge, A. N. *et al.* (1997) *Cell* 89, 115–126
59 Hicke, L. and Riezman, H. (1996) *Cell* 84, 277–287
60 Kölling, R. and Losko, S. (1997) *EMBO J.* 16, 2251–2261
61 Horak, J. and Wolf, D. H. (1997) *J. Bacteriol.* 179, 1541–1549
62 Strous, G. J. *et al.* (1996) *EMBO J.* 15, 3806–3812
63 Hiller, M. M., Finger, A., Schweiger, M. and Wolf, D. H. (1996) *Science* 273, 1725–1728
64 Wiertz, E. J. H. J. *et al.* (1996) *Nature* 384, 432–438
65 Werner, E. D., Brodsky, J. L. and McCracken, A. A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 13797–13801
66 Lopito, R. R. (1997) *Cell* 88, 427–430
67 Lévy, F., Johnsson, N., Rümenapf, T. and Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 4907–4912
68 Webb, E. C., ed. (1992) *Enzyme Nomenclature*, p. 527, Academic Press
69 Finley, D., Özkaynak, E. and Varshavsky, A. (1987) *Cell* 48, 1035–1046
70 McGrath, J. P., Jentsch, S. and Varshavsky, A. (1991) *EMBO J.* 10, 227–236